MATHEMATISCHE ANNALEN

138. BAND

MATHEMATISCHE ANNALEN

BEGRÜNDET 1868 DURCH

ALFRED CLEBSCH UND CARL NEUMANN

FORTGEFÜHRT DURCH

FELIX KLEIN — DAVID HILBERT

OTTO BLUMENTHAL — ERICH HECKE

GEGENWÄRTIG HERAUSGEGEBEN VON

138. BAND

SPRINGER-VERLAG
BERLIN · GÖTTINGEN · HEIDELBERG
1959

Unveränderter Nachdruck 1971
Springer-Verlag Berlin Heidelberg New York

Springer-Verlag OHG/Berlin · Göttingen · Heidelberg

Printed in Germany

Inhalt des 138. Bandes

Huber, H.
Math. Annalen 138, 1—26 (1959)

Zur analytischen Theorie hyperbolischer Raumformen und Bewegungsgruppen

Von
Heinz Huber in Basel

1. Probleme und Ergebnisse

1.1. Es sei $\mathfrak{F}$ eine zweidimensionale, orientierbare und geschlossene analytische Riemannsche Mannigfaltigkeit mit konstanter Gaußscher Krümmung -1, also eine hyperbolische Raumform im Sinne von Felix Klein.

Unter einem geschlossenen Weg $\mathfrak{w}$ auf $\mathfrak{F}$ verstehen wir eine stetige Abbildung

$$\mathfrak{w} : p(t)\,, \quad 0 \leq t \leq 1\,, \quad p(0) = p(1)$$

des Einheitsintervalles in die Fläche $\mathfrak{F}$. Zwei geschlossene Wege heißen homotop, wenn sie auf $\mathfrak{F}$ stetig ineinander deformierbar sind. Dieser Homotopiebegriff ist ein Äquivalenzbegriff in der Menge aller geschlossenen Wege auf $\mathfrak{F}$ und bewirkt daher eine Einteilung dieser Menge in Homotopieklassen $\mathfrak{W}$. Es sei insbesondere $\mathfrak{O}$ die Klasse der in einen Punkt deformierbaren Wege, und $\mathfrak{W}^n$, $n \geq 1$, diejenige Homotopieklasse, welche die n-fach durchlaufenen Wege der Klasse $\mathfrak{W}$ enthält. Eine Klasse $\mathfrak{P}$ heiße primitiv, wenn sie nicht als Potenz $\mathfrak{Q}^n$, $n > 1$, darstellbar ist. Wir werden übrigens sehen (5.5), daß jede Klasse $\mathfrak{W} \neq \mathfrak{O}$ genau eine Darstellung $\mathfrak{W} = \mathfrak{P}^n$ besitzt, in welcher $n \geq 1$ und $\mathfrak{P}$ primitiv ist. Die dadurch eindeutig bestimmte natürliche Zahl $\nu(\mathfrak{W}) = n$ heiße die Vielfachheit der Klasse $\mathfrak{W}$.

Jede Homotopieklasse $\mathfrak{W}$ enthält Wege $\mathfrak{w}$ von endlicher Länge $\int\limits_{\mathfrak{w}} ds$. Wir definieren nun die Länge $\mu(\mathfrak{W})$ einer Klasse $\mathfrak{W}$ durch

$$\mu(\mathfrak{W}) = \inf_{\mathfrak{w} \in \mathfrak{W}} \int\limits_{\mathfrak{w}} ds\,.$$

In der vorliegenden Arbeit stellen wir uns die Aufgabe, die asymptotische Verteilung der Längen $\mu(\mathfrak{W})$ und $\mu(\mathfrak{P})$ zu studieren. Im Verlaufe dieses Studiums werden sich interessante Beziehungen unserer Aufgabe zu einem gewissen Gitterpunktproblem in der hyperbolischen Ebene und zum Laplace-Beltramischen Eigenwertproblem $\Delta\,\varphi + \lambda\,\varphi = 0$ auf $\mathfrak{F}$ ergeben.

1.2. Zur Lösung unserer Aufgabe bedienen wir uns mit Vorteil des engen Zusammenhanges, der bekanntlich zwischen den hyperbolischen Raumformen und den diskontinuierlichen Bewegungsgruppen der hyperbolischen Ebene besteht. Sei $\mathfrak{H}$ universelle Überlagerungsfläche von $\mathfrak{F}$; wir denken uns die differentialgeometrische Struktur von $\mathfrak{F}$ auf $\mathfrak{H}$ durchgedrückt, so daß jeder Überlagerungsweg auf $\mathfrak{H}$ dieselbe Länge besitzt wie sein Grundweg auf $\mathfrak{F}$.

Dann ist $\mathfrak{H}$ ersichtlich eine einfach zusammenhängende, vollständige analytische Fläche konstanter Krümmung -1, also eine hyperbolische Ebene. Die zugehörige Decktransformationsgruppe Λ ist eine diskontinuierliche Gruppe von (eigentlichen) Bewegungen von $\mathfrak{H}$; da $\mathfrak{F}$ geschlossen ist, besitzt sie einen kompakten meßbaren Fundamentalbereich[1]).

Wir ordnen jeder Homotopieklasse $\mathfrak{W}$ eine Klasse $\Phi_{\mathfrak{W}}$ konjugierter Elemente von Λ in folgender Weise zu: Ist

$$\mathfrak{w} : p(t)\,, \quad 0 \leq t \leq 1\,, \quad p(0) = p(1)$$

ein Weg der Klasse $\mathfrak{W}$ und $\bar{\mathfrak{w}}$ irgendeiner seiner Überlagerungswege auf $\mathfrak{H}$, so gibt es genau eine Bewegung $T \in \Lambda$, welche den Anfangspunkt von $\bar{\mathfrak{w}}$ in den Endpunkt von $\bar{\mathfrak{w}}$ überführt. Dann sei

$$\Phi_{\mathfrak{W}} = \{STS^{-1} | S \in \Lambda\}\,.$$

Die so definierte Klasse $\Phi_{\mathfrak{W}}$ ist unabhängig von der Wahl des Überlagerungsweges $\bar{\mathfrak{w}}$ von $\mathfrak{w}$ und unabhängig von der Wahl des Repräsentanten $\mathfrak{w} \in \mathfrak{W}$. Die Zuordnung $\mathfrak{W} \to \Phi_{\mathfrak{W}}$ bildet die Menge aller Homotopieklassen $\mathfrak{W}$ umkehrbar eindeutig auf die Menge aller Klassen Φ konjugierter Elemente von Λ ab; insbesondere gilt $\Phi_{\mathfrak{O}} = \{E\}$, wenn E das Einselement von Λ ist[2]). Man erkennt nun leicht, daß für jedes $T \in \Phi_{\mathfrak{W}}$

$$\inf_{p \in \mathfrak{H}} \varrho(Tp, p) = \mu(\mathfrak{W})\,;$$

dabei ist $\varrho(p, q)$ die Distanz der Punkte $p, q \in \mathfrak{H}$, also die untere Grenze der Längen der Wege, die auf $\mathfrak{H}$ von p nach q führen. Somit gilt, wenn wir die Verschiebungslänge einer Bewegung T von $\mathfrak{H}$ durch

$$\mu(T) = \inf_{p \in \mathfrak{H}} \varrho(Tp, p) \tag{1}$$

definieren,

$$\mu(\mathfrak{W}) = \mu(T) \quad \text{für alle } T \in \Phi_{\mathfrak{W}}\,. \tag{2}$$

Weil $\mathfrak{F}$ geschlossen ist, ist natürlich $\mu(\mathfrak{W}) > 0$ für $\mathfrak{W} \neq \mathfrak{O}$. Daher ist $\mu(T) > 0$ für alle $T \in \Lambda - E$. Die einzigen Bewegungen von $\mathfrak{H}$ mit positiver Verschiebungslänge sind aber die hyperbolischen Translationen[3]). Somit ist Λ eine diskontinuierliche Translationsgruppe.

1.3. Die Gleichungen (1), (2) legen es einigermaßen nahe, zuerst die asymptotische Verteilung der Distanzen $\varrho(Tp, p)$ oder, allgemeiner, die Verteilung der Distanzen $\varrho(Tp, q)$, $T \in \Lambda$, zu untersuchen. Dazu betrachten wir die Dirichlet-Reihe

$$G_{\mathfrak{F}}(p, q; s) = \sum_{T \in \Lambda} \operatorname{Cos}^{-s} \varrho(Tp, q),\ s = \sigma + it,\ p, q \in \mathfrak{H}\,, \tag{3}$$

welche für $\sigma > 1$ absolut konvergiert (3.7). Wir werden in 3.8 und 3.9 beweisen:

Satz 1. *Für alle $s = \sigma + it$, $\sigma > 1$, und $q \in \mathfrak{H}$ gilt:*

(α) *$G_{\mathfrak{F}}(p, q; s)$ ist eine auf $\mathfrak{H}$ stetige und bezüglich Λ automorphe Funktion von p.*

[1]) Vgl. z. B. [8], pag. 148—149.
[2]) Vgl. hierzu [7], § 49.
[3]) Vgl. Abschnitt 2.8.

(b) $G_{\mathfrak{F}}(p, q; s)$ *ist zweimal stetig differenzierbar nach lokalen regulären Koordinaten von p und erfüllt die Funktionalgleichung*

$$\Delta_p G_{\mathfrak{F}}(p, q; s) + s(1 - s)\, G_{\mathfrak{F}}(p, q; s) + s(1 + s)\, G_{\mathfrak{F}}(p, q; s + 2) = 0\,,$$

wenn Δ *der zu* $\mathfrak{F}$ *und* $\mathfrak{H}$ *gehörige Laplace-Beltrami-Operator ist.*

1.4. Dieser Satz führt uns dazu, das Eigenwertproblem $\Delta\varphi + \lambda\varphi = 0$ auf $\mathfrak{F}$ zu betrachten: Die Zahl λ heiße *Eigenwert*, wenn es auf $\mathfrak{F}$ eine nicht überall verschwindende Funktion φ gibt, welche (nach lokalen regulären Koordinaten) zweimal stetig differenzierbar ist und die Differentialgleichung $\Delta\varphi + \lambda\varphi = 0$ erfüllt. Jede solche Funktion heiße eine zum Eigenwert λ gehörige *Eigenfunktion*. Da $\mathfrak{F}$ eine geschlossene analytische Mannigfaltigkeit ist, läßt sich dieses Eigenwertproblem im Rahmen klassischer Theorien behandeln[4]). Es ergeben sich bekanntlich folgende Tatsachen: Es gibt unendlich viele Eigenwerte; sie sind, abgesehen vom trivialen Eigenwert $\lambda = 0$, positiv und häufen sich nirgends im Endlichen. Wir können sie daher in folgender Weise numerieren:

$$\lambda_0 = 0 < \lambda_1 < \lambda_2 < \cdots < \lambda_n < \cdots, \quad \lim_{n\to\infty} \lambda_n = +\infty\,. \tag{4}$$

Der lineare Raum der zum Eigenwert λ_n gehörigen Eigenfunktionen besitzt eine endliche Dimension $m_n \geqq 1$; insbesondere ist

$$m_0 = 1\,, \tag{5}$$

weil jede zum Eigenwert $\lambda_0 = 0$ gehörige Eigenfunktion offenbar harmonisch auf der geschlossenen Fläche $\mathfrak{F}$ und somit konstant ist. Die für alle $\lambda \geqq 0$ definierte Funktion

$$m_{\mathfrak{F}}(\lambda) = \begin{cases} m_n & \text{für } \lambda = \lambda_n \\ 0 & \text{für } \lambda \notin \{\lambda_n\} \end{cases} \tag{6}$$

nennen wir das *Eigenwertspektrum* von $\mathfrak{F}$. Unter der *Spektralfolge* $\{\lambda'_n | n \geqq 0\}$ verstehen wir diejenige nicht abnehmende Zahlenfolge, in der jeder Eigenwert λ_n genau m_n-mal auftritt. Offenbar gilt:

$$\lambda'_0 = 0 < \lambda'_1 = \lambda_1 \leqq \lambda'_2 \leqq \cdots \leqq \lambda'_n \leqq \cdots, \quad \lim_{n\to\infty} \lambda'_n = +\infty\,. \tag{7}$$

Wir führen gleich noch eine weitere mit den Eigenwerten verknüpfte Folge ein, die bald eine wichtige Rolle spielen wird: Definieren wir

$$\begin{aligned} s^+(\lambda) &= \frac{1}{2} + \frac{1}{2}\sqrt[+]{1 - 4\lambda}, & s^-(\lambda) &= \frac{1}{2} - \frac{1}{2}\sqrt[+]{1 - 4\lambda} & \text{für } & 0 \leqq \lambda \leqq \frac{1}{4} \\ s^+(\lambda) &= \frac{1}{2} + \frac{i}{2}\sqrt[+]{4\lambda - 1}, & s^-(\lambda) &= \frac{1}{2} - \frac{i}{2}\sqrt[+]{4\lambda - 1} & \text{für } & \frac{1}{4} < \lambda < +\infty, \end{aligned} \tag{8}$$

so gilt

$$s^+(\lambda_0) = s^+(\lambda'_0) = 1\,, \quad s^-(\lambda_0) = s^-(\lambda'_0) = 0\,, \tag{9}$$

und die Punkte $s^+(\lambda_n)$, $s^-(\lambda_n)$ liegen symmetrisch in bezug auf den Punkt $s = \frac{1}{2}$ der komplexen $s = \sigma + it$-Ebene. Wenn $\lambda_1 \geqq \frac{1}{4}$, so liegt die Punktmenge

[4]) Vgl. z. B. [3] und insbesondere [6].

$\{s^+(\lambda_n), s^-(\lambda_n) | n \geqq 1\}$ auf der Geraden $\sigma = \frac{1}{2}$; wenn dagegen $0 < \lambda_1 < \frac{1}{4}$, so liegen endlich viele Punkte dieser Menge im offenen Intervall $0 < s < 1$, alle übrigen wieder auf der Geraden $\sigma = \frac{1}{2}$. Es gibt daher stets ein solches $\delta = \delta(\lambda_1)$, $0 < \delta \leqq \frac{1}{2}$, daß die Menge $\{s^+(\lambda_n), s^-(\lambda_n) | n \geqq 1\}$ in der Halbebene $\sigma \leqq 1 - \delta$ liegt.

1.5. Das asymptotische Verteilungsgesetz der Spektralfolge $\{\lambda_n'\}$ ist seit den Untersuchungen von H. Weyl und anderen wohlbekannt[5]):

$$\sum_{\lambda_n' \leq t} 1 = \sum_{\lambda \leq t} m_{\mathfrak{F}}(\lambda) \sim \frac{t}{4\pi} \iint_{\mathfrak{F}} d\omega \qquad \text{für} \quad t \to +\infty .$$

Da $\mathfrak{F}$ die konstante Krümmung -1 besitzt, so ist nach dem Gaußschen Satz von der Curvatura integra

$$\iint_{\mathfrak{F}} d\omega = 4\pi(g_{\mathfrak{F}} - 1)\,, \tag{10}$$

wenn $g_{\mathfrak{F}} > 1$ das Geschlecht der orientierbaren geschlossenen Fläche $\mathfrak{F}$ ist. Wir haben somit

$$g_{\mathfrak{F}} - 1 = \lim_{t \to \infty} \frac{1}{t} \sum_{\lambda \leq t} m_{\mathfrak{F}}(\lambda)\,. \tag{11}$$

1.6. Da die Eigenwerte reell sind, gibt es ein zur Spektralfolge $\{\lambda_n'\}$ gehöriges Orthonormalsystem $\{\varphi_n\}$ *reeller* Eigenfunktionen. Wir dürfen und wollen diese Eigenfunktionen als Funktionen auf der universellen Überlagerungsfläche $\mathfrak{H}$ von $\mathfrak{F}$ auffassen. Dann ist $\varphi_n(p)$, $n \geqq 0$, eine auf $\mathfrak{H}$ zweimal stetig differenzierbare und bezüglich Λ automorphe Lösung der Differentialgleichung $\Delta\varphi_n + \lambda_n' \varphi_n = 0$, und es gilt

$$\iint_{\mathfrak{B}} \varphi_m(p)\, \varphi_n(p)\, d\omega_p = \begin{cases} 1 & \text{für } m = n \\ 0 & \text{für } m \neq n \end{cases}, \tag{12}$$

wenn $d\omega_p$ das Flächenelement von $\mathfrak{H}$ und $\mathfrak{B}$ irgendein meßbarer Fundamentalbereich der Translationsgruppe Λ ist. Da die normierte Eigenfunktion $\varphi_0(p)$ konstant ist, ergibt sich

$$\varphi_0(p) \equiv \left(\iint_{\mathfrak{B}} d\omega_p\right)^{-\frac{1}{2}} = \left(\iint_{\mathfrak{F}} d\omega\right)^{-\frac{1}{2}},$$

also wegen (10)

$$\varphi_0(p) \equiv (4\pi(g_{\mathfrak{F}} - 1))^{-\frac{1}{2}}. \tag{13}$$

1.7. Das Orthonormalsystem $\{\varphi_n\}$ hat nun bekanntlich folgende Eigenschaften:

(a) $\sum_{n=1}^{\infty} \left(\frac{1}{\lambda_n'}\right)^2$ konvergiert, und $\sum_{n=1}^{\infty} \left(\frac{\varphi_n(p)}{\lambda_n'}\right)^2$ konvergiert gleichmäßig auf $\mathfrak{H}$.

[5]) Vgl. z. B. [6].

(b) Für jede auf $\mathfrak{H}$ zweimal stetig differenzierbare und bezüglich Λ automorphe Funktion $f(p)$ gilt

$$f(p) = \sum_{n=0}^{\infty} a_n \varphi_n(p), \quad a_n = \iint_{\mathfrak{F}} f(p)\, \varphi_n(p)\, d\omega_p .$$

1.8. Daher können wir jetzt auf Grund von Satz 1 die Funktion $G_{\mathfrak{F}}(p,q;s)$ für festes q und s in eine Fourier-Reihe nach den Eigenfunktionen $\varphi_n(p)$ entwickeln. Dies gelingt dank der Funktionalgleichung von $G_{\mathfrak{F}}(p,q;s)$ in recht expliziter Weise; wir werden schließlich (4.17, 4.21) zu folgendem Ergebnis geführt:

Satz 2. *(a) Die Reihe*

$$H_{\mathfrak{F}}(p,q;s) = \sum_{n=1}^{\infty} \Gamma\left(\frac{s-s^+(\lambda_n)}{2}\right) \Gamma\left(\frac{s-s^-(\lambda_n)}{2}\right) \varphi_n(p)\, \varphi_n(q)$$

konvergiert absolut für $s \notin P = \{s^+(\lambda_n) - 2l, s^-(\lambda_n) - 2l \mid n \geq 1, l \geq 0\}$ und stellt eine in der ganzen s-Ebene meromorphe Funktion dar, deren Pole in der Punktmenge P enthalten sind. $H_{\mathfrak{F}}(p,q;s)$ besitzt im Punkte $s^{\pm}(\lambda_n) - 2l \in P$ denselben Hauptteil wie die Funktion

$$\Gamma\left(\frac{s-s^+(\lambda_n)}{2}\right) \Gamma\left(\frac{s-s^-(\lambda_n)}{2}\right) \sum_{\lambda_m = \lambda_n} \varphi_m(p)\, \varphi_m(q) .$$

(b) In der Halbebene $\sigma > 1$ gilt

$$G_{\mathfrak{F}}(p,q;s) = \frac{1}{2(g_{\mathfrak{F}}-1)} \frac{1}{s-1} + \sqrt{\pi}\, \frac{2^{s-1}}{\Gamma(s)} H_{\mathfrak{F}}(p,q;s) .$$

Damit ist es uns gelungen, die nur in der Halbebene $\sigma > 1$ konvergente Dirichlet-Reihe $G_{\mathfrak{F}}(p,q;s)$ zu einer in der ganzen s-Ebene meromorphen Funktion analytisch fortzusetzen. Da P offenbar in der Halbebene $\sigma \leq 1 - \delta$, $0 < \delta(\lambda_1) \leq \frac{1}{2}$ liegt, so ergibt sich aus Satz 2: In der Halbebene $\sigma > 1$ gilt

$$\sum_{T \in \Lambda} \mathrm{Cos}^{-s} \varrho(Tp, q) = \frac{1}{2(g_{\mathfrak{F}}-1)} \frac{1}{s-1} + R(p,q;s) ,$$

wobei die Funktion $R(p,q;s)$ sogar in der Halbebene $\sigma > 1 - \delta$ regulär analytisch ist. Hieraus folgt aber nach dem Tauberschen Theorem von WIENER-IKEHARA[6]) sofort der

Satz 3. *Es sei $N(p,q,t)$ die Anzahl der Elemente der Menge*

$$\{T \mid T \in \Lambda, \varrho(Tp,q) \leq t\} .$$

Dann gilt für $t \to +\infty$

$$N(p,q,t) \sim \frac{1}{4(g_{\mathfrak{F}}-1)} e^t .$$

In Analogie zum Fall einer diskontinuierlichen Translationsgruppe mit kompaktem Fundamentalbereich der euklidischen Ebene nennen wir die Punktmenge $\mathfrak{G}_p = \{Tp \mid T \in \Lambda\}$ ein hyperbolisches Gitter. $N(p,q,t)$ ist dann

[6]) [9], pag. 44.

nichts anderes als die Anzahl der Punkte des Gitters $\mathfrak{G}_p$, welche in der hyperbolischen Kreisscheibe mit dem Zentrum q und dem Radius t liegen. Satz 3 liefert für gewisse arithmetisch erzeugte Translationsgruppen bemerkenswerte zahlentheoretische Ergebnisse, auf die in einer späteren Arbeit eingegangen werden soll.

1.9. Auf Grund der Entwicklung von $G_{\mathfrak{F}}(p, q; s)$ nach den Eigenfunktionen $\varphi_n(p)$ werden wir weiter zeigen (4.18, 4.22):

Satz 4. (a) *Die Reihe*

$$H_{\mathfrak{F}}(s) = \sum_{n=1}^{\infty} m_n \, \Gamma\left(\frac{s - s^+(\lambda_n)}{2}\right) \Gamma\left(\frac{s - s^-(\lambda_n)}{2}\right) \tag{14}$$

konvergiert absolut für $s \notin P$ *und stellt eine in der ganzen s-Ebene meromorphe Funktion mit der Polmenge* P *dar.* $H_{\mathfrak{F}}(s)$ *besitzt im Punkte* $s^{\pm}(\lambda_n) - 2l \in P$ *denselben Hauptteil wie die Funktion* $m_n \, \Gamma\left(\frac{s - s^+(\lambda_n)}{2}\right) \Gamma\left(\frac{s - s^-(\lambda_n)}{2}\right)$.

(b) *In der Halbebene* $\sigma > 1$ *gilt*

$$\iint_{\mathfrak{B}} G_{\mathfrak{F}}(p, p; s) \, d\omega_p = \frac{2\pi}{s-1} + \sqrt{\pi} \, \frac{2^{s-1}}{\Gamma(s)} \, H_{\mathfrak{F}}(s) \, .$$

Andererseits werden wir aber durch gliedweise Integration der Dirichlet-Reihe $G_{\mathfrak{F}}(p, p; s)$ zeigen (5.14):

Satz 5. *Die Dirichlet-Reihe*

$$D_{\mathfrak{F}}(s) = \sum_{\mathfrak{W} \neq \mathfrak{O}} \frac{\mu(\mathfrak{W})}{\nu(\mathfrak{W})} \left(\frac{\operatorname{Cos} \mu(\mathfrak{W})}{\operatorname{Cos} \mu(\mathfrak{W}) - 1}\right)^{\frac{1}{2}} \operatorname{Cos}^{-s} \mu(\mathfrak{W}) \tag{15}$$

konvergiert absolut für $\sigma > 1$, *und es gilt*

$$D_{\mathfrak{F}}(s) = \frac{1}{\sqrt{\pi}} \, \frac{\Gamma(s)}{\Gamma(s - \frac{1}{2})} \iint_{\mathfrak{B}} G_{\mathfrak{F}}(p, p; s) \, d\omega_p - 4\sqrt{\pi} \, (g_{\mathfrak{F}} - 1) \, \frac{\Gamma(s)}{\Gamma(s - \frac{1}{2})} \, .$$

Aus Satz 4 und Satz 5 ergibt sich nun der

Satz 6. *Die Funktion*

$$M_{\mathfrak{F}}(s) = \sqrt{\pi} \, \frac{\Gamma(s)}{\Gamma(s - \frac{1}{2})} \left(\frac{2}{s-1} - 4(g_{\mathfrak{F}} - 1)\right) + \frac{2^{s-1}}{\Gamma(s - \frac{1}{2})} \, H_{\mathfrak{F}}(s) \tag{16}$$

ist meromorph in der ganzen s-Ebene, und es gilt in der Halbebene $\sigma > 1$

$$D_{\mathfrak{F}}(s) = M_{\mathfrak{F}}(s) \, . \tag{17}$$

Wir haben somit die nur für $\sigma > 1$ konvergente Dirichlet-Reihe $D_{\mathfrak{F}}(s)$ zu einer in der ganzen s-Ebene meromorphen Funktion analytisch fortgesetzt. In den folgenden Nummern wollen wir die Folgerungen aus diesen Ergebnissen ziehen.

1.10. Zunächst ziehen wir einige einfache Schlüsse aus der Tatsache, daß die in der Halbebene $\sigma > 1$ konvergente Dirichlet-Reihe $D_{\mathfrak{F}}(s)$ positive Koeffizienten besitzt. Da $\bar{\nu}(\mathfrak{P}) = 1$ für primitive Klassen $\mathfrak{P}$, so folgt aus (15)

$$\sum_{\mathfrak{P}} \mu(\mathfrak{P}) \left(\frac{\operatorname{Cos} \mu(\mathfrak{P})}{\operatorname{Cos} \mu(\mathfrak{P}) - 1}\right)^{\frac{1}{2}} \operatorname{Cos}^{-\sigma} \mu(\mathfrak{P}) < D_{\mathfrak{F}}(\sigma) < +\infty \quad \text{für } \sigma > 1 \, .$$

Daraus ergibt sich offensichtlich

(18) $$\pi_{\mathfrak{F}}(t) = \sum_{\mu(\mathfrak{P}) \leq t} 1 < \infty \quad \text{für } t < \infty .$$

Nach 1.1 besitzt jede Homotopieklasse $\mathfrak{W} \neq \mathfrak{O}$ eine eindeutige Darstellung $\mathfrak{W} = \mathfrak{P}^n$, in welcher $\mathfrak{P}$ primitiv und $n \geqq 1$. Nun gilt aber[7])

(19) $$\mu(\mathfrak{P}^n) = n\,\mu(\mathfrak{P}) .$$

Somit folgt aus (18)

(20) $$\omega_{\mathfrak{F}}(t) = \sum_{\mu(\mathfrak{W}) \leq t} 1 < \infty \quad \text{für } t < \infty .$$

Die Zahlenmenge $\{\mu(\mathfrak{W}) \mid \mathfrak{W} \neq \mathfrak{O}\}$ kann daher in eine monoton nach unendlich wachsende Folge angeordnet werden:

(21) $$\{\mu(\mathfrak{W}) | \mathfrak{W} \neq \mathfrak{O}\} = \{\mu_n | n \geqq 0\}, 0 < \mu_0 < \mu_1 < \cdots < \mu_n < \cdots; \lim_{n\to\infty} \mu_n = +\infty .$$

Zu jedem μ_n gibt es dann auf $\mathfrak{F}$ mindestens eine, aber höchstens endlich viele Homotopieklassen $\mathfrak{W}$ mit $\mu(\mathfrak{W}) = \mu_n$. Wir ordnen nun jedem μ_n das Gewicht

$$h_n = \sum_{\mu(\mathfrak{W}) = \mu_n} 1/\nu(\mathfrak{W})$$

zu. Die für alle $\mu \geqq 0$ definierte Funktion

(22) $$h_{\mathfrak{F}}(\mu) = \begin{cases} h_n & \text{für } \mu = \mu_n \\ 0 & \text{für } \mu \notin \{\mu_n\} \end{cases}$$

heiße das *Längenspektrum* von $\mathfrak{F}$. Dieses Längenspektrum beschreibt offenbar die Maßverhältnisse im Großen auf $\mathfrak{F}$. Die Dirichlet-Reihe (15) kann jetzt auch folgendermaßen geschrieben werden:

(23) $$D_{\mathfrak{F}}(s) = \sum_{n=0}^{\infty} h_n \mu_n \left(\frac{\operatorname{Cos} \mu_n}{\operatorname{Cos} \mu_n - 1}\right)^{\frac{1}{2}} \operatorname{Cos}^{-s} \mu_n .$$

Wegen der absoluten Konvergenz dieser Reihe für $\sigma > 1$ konvergiert

$$D'_{\mathfrak{F}}\left(\frac{3}{2} + it\right) = \sum_{n=0}^{\infty} \frac{h_n \mu_n}{\operatorname{Cos} \mu_n \sqrt{\operatorname{Cos} \mu_n - 1}}\, e^{-it \log \operatorname{Cos} \mu_n}$$

gleichmäßig in $-\infty < t < +\infty$. Hieraus und aus (21), (22) folgt nach einem Satz aus der Theorie der fastperiodischen Funktionen[8])

$$\lim_{t\to+\infty} \frac{1}{t} \int_0^t D_{\mathfrak{F}}\left(\frac{3}{2} + i\tau\right) e^{i\tau \log \operatorname{Cos} \mu}\, d\tau = \frac{\mu\, h_{\mathfrak{F}}(\mu)}{\sqrt{\operatorname{Cos} \mu - 1}\ \operatorname{Cos} \mu} \quad \text{für } \mu \geqq 0 .$$

Somit folgt aus Satz 6

(24) $$h_{\mathfrak{F}}(\mu) = \frac{1}{\mu} \sqrt{\operatorname{Cos} \mu - 1}\ \operatorname{Cos} \mu \cdot \lim_{t\to+\infty} \frac{1}{t} \int_0^t M_{\mathfrak{F}}\left(\frac{3}{2} + i\tau\right) e^{i\tau \log \operatorname{Cos} \mu}\, d\tau .$$

---

[7]) Wenn $T \in \Phi_{\mathfrak{P}}$, so ist offensichtlich $T^n \in \Phi_{\mathfrak{P}^n}$ und daher nach 1.2 (2): $\mu(\mathfrak{P}) = \mu(T)$, $\mu(\mathfrak{P}^n) = \mu(T^n)$; nach 2.6 ist aber $\mu(T^n) = n\,\mu(T)$ und somit $\mu(\mathfrak{P}^n) = n\,\mu(\mathfrak{P})$.

[8]) [1], pag. 51, Satz XII.

Damit haben wir das Längenspektrum mit Hilfe der meromorphen Funktion $M_{\mathfrak{F}}(s)$ dargestellt, was alsbald von Bedeutung sein wird. Wir wollen nun zeigen, daß auch das Eigenwertspektrum und das Geschlecht von $\mathfrak{F}$ aus der Funktion $M_{\mathfrak{F}}(s)$ berechnet werden kann. In der Tat ergibt sich aus (16) und Satz 4 (a) unter Berücksichtigung von (5), (8), (9) durch eine leichte Rechnung

$$(25)\qquad m_n = \begin{cases} (\operatorname{Res} 2^{-s} M_{\mathfrak{F}}(s))_{s=s^+(\lambda_n)} & \text{für } \lambda_n \neq \frac{1}{4} \\ \frac{1}{2}(\operatorname{Res} 2^{-s} M_{\mathfrak{F}}(s))_{s=s^+(\lambda_n)} & \text{für } \lambda_n = \frac{1}{4}\,. \end{cases}$$

Da der Punkt $s^+(\lambda)$ für $\lambda \geq 0$, $\lambda \notin \{\lambda_n\}$ offenbar eine reguläre Stelle von $M_{\mathfrak{F}}(s)$ ist, so folgt aus (6) und (25)

$$(26)\qquad m_{\mathfrak{F}}(\lambda) = \left\{\begin{array}{ll} (\operatorname{Res} 2^{-s} M_{\mathfrak{F}}(s))_{s=s^+(\lambda)} & \text{für } \lambda \neq \frac{1}{4} \\ \frac{1}{2}(\operatorname{Res} 2^{-s} M_{\mathfrak{F}}(s))_{s=s^+(\lambda)} & \text{für } \lambda = \frac{1}{4} \end{array}\right\},\ \lambda \geq 0\,.$$

Nach Satz 4(a) ist $s = 0$ eine reguläre Stelle von $H_{\mathfrak{F}}(s)$. Daher berechnet man leicht aus (16)

$$(27)\qquad 2g_{\mathfrak{F}} - 1 = (\operatorname{Res} M_{\mathfrak{F}}(s))_{s=0}\,.$$

1.11. Es seien nun $\mathfrak{F}_1, \mathfrak{F}_2$ geschlossene Raumformen mit gleichem Längenspektrum: $h_{\mathfrak{F}_1}(\mu) = h_{\mathfrak{F}_2}(\mu)$ für $\mu \geq 0$. Dann folgt aus (22) und (23): $D_{\mathfrak{F}_1}(s) = D_{\mathfrak{F}_2}(s)$ für $\sigma > 1$. Daher stimmen nach Satz 6 die beiden meromorphen Funktionen $M_{\mathfrak{F}_1}(s)$, $M_{\mathfrak{F}_2}(s)$ in der Halbebene $\sigma > 1$ überein und sind somit in der ganzen s-Ebene identisch. Daraus folgt aber nach (26) und (27): $m_{\mathfrak{F}_1}(\lambda) = m_{\mathfrak{F}_2}(\lambda)$, $g_{\mathfrak{F}_1} = g_{\mathfrak{F}_2}$. Wir haben somit den

Satz 7. *Geschlossene hyperbolische Raumformen mit gleichem Längenspektrum besitzen auch gleiches Eigenwertspektrum und gleiches Geschlecht.*

Es seien jetzt $\mathfrak{F}_1, \mathfrak{F}_2$ Raumformen mit gleichem Eigenwertspektrum: $m_{\mathfrak{F}_1}(\lambda) = m_{\mathfrak{F}_2}(\lambda)$ für $\lambda \geq 0$. Dann folgt aus (6) und (14)

$$(28)\qquad H_{\mathfrak{F}_1}(s) = H_{\mathfrak{F}_2}(s)\,.$$

Ferner folgt aus (11)

$$(29)\qquad g_{\mathfrak{F}_1} = g_{\mathfrak{F}_2}\,.$$

Aus (28) und (29) folgt nun nach (16): $M_{\mathfrak{F}_1}(s) = M_{\mathfrak{F}_2}(s)$. Somit ist nach (24) $h_{\mathfrak{F}_1}(\mu) = h_{\mathfrak{F}_2}(\mu)$ für $\mu \geq 0$. Wir haben also den

Satz 8. *Geschlossene Raumformen mit gleichem Eigenwertspektrum besitzen auch gleiches Längenspektrum und gleiches Geschlecht.*

1.12. Nach Satz 4(a) liegen sämtliche Pole von $H_{\mathfrak{F}}(s)$ in der Halbebene $\sigma \leqq 1 - \delta$, $0 < \delta(\lambda_1) \leqq \frac{1}{2}$. Daher ergibt sich aus Satz 6: In der Halbebene $\sigma > 1$ gilt:

$$D_{\mathfrak{F}}(s) = \sum_{\mathfrak{W} \neq \mathfrak{O}} \frac{\mu(\mathfrak{W})}{\nu(\mathfrak{W})} \left(\frac{\operatorname{Cos}\mu(\mathfrak{W})}{\operatorname{Cos}\mu(\mathfrak{W}) - 1}\right)^{\frac{1}{2}} \operatorname{Cos}^{-s}\mu(\mathfrak{W}) = \frac{2}{s-1} + R(s)\,,$$

wobei die Funktion $R(s)$ sogar in der Halbebene $\sigma > 1 - \delta$ regulär analytisch

ist. Daraus folgt aber nach dem Tauberschen Theorem von WIENER-IKEHARA[9]) sofort

$$\sum_{0<\mu(\mathfrak{W})\leq t} \frac{\mu(\mathfrak{W})}{\nu(\mathfrak{W})} \left(\frac{\operatorname{Cos}\mu(\mathfrak{W})}{\operatorname{Cos}\mu(\mathfrak{W})-1}\right)^{\frac{1}{2}} \sim e^t \quad \text{für } t \to +\infty .$$

Hieraus folgt wegen $\left(\frac{\operatorname{Cos}\mu}{\operatorname{Cos}\mu - 1}\right)^{\frac{1}{2}} \to 1$ für $\mu \to +\infty$

$$\psi(t) = \sum_{0<\mu(\mathfrak{W})\leq t} \frac{\mu(\mathfrak{W})}{\nu(\mathfrak{W})} \sim e^t \quad \text{für } t \to +\infty . \tag{30}$$

Um daraus eine asymptotische Aussage über

$$\pi(t) = \sum_{\mu(\mathfrak{P})\leq t} 1 \tag{31}$$

herzuleiten, gehen wir ähnlich vor wie beim Beweis des Primzahlsatzes. Es sei

$$\vartheta(t) = \sum_{\mu(\mathfrak{P})\leq t} \mu(\mathfrak{P}) . \tag{32}$$

Nach 1.1 besitzt jede Klasse $\mathfrak{W} \neq \mathfrak{O}$ eine eindeutige Darstellung $\mathfrak{W} = \mathfrak{P}^n$, in welcher $\mathfrak{P}$ primitiv und $n \geqq 1$. Es gilt aber $\nu(\mathfrak{P}^n) = n$ und $\mu(\mathfrak{P}^n) = n\mu(\mathfrak{P})$[10]). Daher ergibt sich aus (30), (32) offensichtlich: $\psi(t) = \vartheta(t) + \vartheta(t/2) + \vartheta(t/3) + \cdots$. Diese Reihe bricht nach endlich vielen Gliedern ab, denn es ist natürlich $\vartheta(t/n) = 0$ sobald $t/n < \mu_0$. Wir haben also

$$\psi(t) = \vartheta(t) + \sum_{n=2}^{N(t)} \vartheta(t/n) , \qquad N(t) = \left[\frac{t}{\mu_0}\right] . \tag{33}$$

Daraus und aus (30) folgt zunächst die Existenz einer endlichen Konstanten $c > 0$ derart, daß $\vartheta(t) \leqq \psi(t) < ce^t$ für $t \geqq 0$. Somit gilt für $t > 2\mu_0$

$$0 \leqq \sum_{n=2}^{N(t)} \vartheta(t/n) < c \sum_{n=2}^{N(t)} e^{t/n} < c \int_1^{t/\mu_0} e^{t/x} dx = ct \int_{\mu_0}^{t} u^{-2} e^u du . \tag{34}$$

Man zeigt aber leicht, daß

$$\int_{\mu_0}^{t} u^{-2} e^u du \sim t^{-2} e^t , \quad t \to +\infty . \tag{35}$$

Somit folgt aus (34): $\sum_{n=2}^{N(t)} \vartheta(t/n) = O(t^{-1}e^t)$. Daher folgt aus (30) und (33)

$$\vartheta(t) \sim e^t \quad \text{für } t \to +\infty . \tag{36}$$

Wegen (31), (32) gilt offensichtlich

$$\pi(t) = \int_{\varepsilon}^{t} u^{-1} d\vartheta(u) , \quad 0 < \varepsilon < \mu_0 .$$

Partielle Integration ergibt

$$\pi(t) - t^{-1}\vartheta(t) = \int_{\mu_0}^{t} u^{-2} \vartheta(u) du .$$

---

[9]) Siehe Anmerkung [6]).

[10]) Vgl. [7]).

Daraus und aus (36), (35) schließt man leicht

$$\pi(t) - t^{-1}\vartheta(t) \sim t^{-2}e^t, \quad t \to +\infty .$$

Somit gilt wegen (36)

Satz 9. *Es sei $\pi_{\mathfrak{F}}(t)$ die Anzahl aller primitiven Homotopieklassen der Länge $\leq t$ auf $\mathfrak{F}$. Dann gilt für $t \to +\infty$*

$$\pi_{\mathfrak{F}}(t) \sim e^t/t .$$

Definieren wir

$$\omega(t) = \sum_{\mu(\mathfrak{B}) \leq t} 1 ,$$

so gilt offenbar $\omega(t) = 1 + \pi(t) + \pi(t/2) + \pi(t/3) + \cdots$, und diese Reihe bricht wiederum ab, sobald $t/n < \mu_0$; wir haben somit

$$\omega(t) = 1 + \pi(t) + \sum_{n=2}^{N(t)} \pi(t/n), \quad N(t) = \left[\frac{t}{\mu_0}\right].$$

Daraus und aus Satz 9 ergibt sich leicht der

Satz 10. *Es sei $\omega_{\mathfrak{F}}(t)$ die Anzahl aller Homotopieklassen der Länge $\leq t$ auf $\mathfrak{F}$. Dann gilt für $t \to +\infty$*

$$\omega_{\mathfrak{F}}(t) \sim e^t/t .$$

Es mag zunächst überraschen, daß das asymptotische Verhalten von $\pi_{\mathfrak{F}}(t)$ und $\omega_{\mathfrak{F}}(t)$ für alle geschlossenen Raumformen $\mathfrak{F}$ gleich ist. Dies ist ein Phänomen ähnlicher Art wie die bekannte Tatsache, daß die Anzahl der Primideale der Norm $\leq t$ in allen algebraischen Zahlkörpern dasselbe asymptotische Verhalten zeigt.

2. Hilfssätze[11])

2.1. Wir erinnern daran, daß $\mathfrak{H}$ eine einfach-zusammenhängende, vollständige analytische Fläche konstanter Krümmung -1 ist. Bekanntlich gibt es auf $\mathfrak{H}$ Systeme $\{x_1(p), x_2(p)\}$ reell-analytischer Funktionen von p mit folgenden Eigenschaften:

(a) $\{x_1(p), x_2(p)\}$ ist ein überall auf $\mathfrak{H}$ reguläres Koordinatensystem, in welchem die metrische Differentialform von $\mathfrak{H}$ die Gestalt

(1) $$ds^2 = x_2^{-2}(dx_1^2 + dx_2^2)$$

besitzt.

(b) Die Zuordnung $p \to z(p) = x_1(p) + i x_2(p)$ bildet $\mathfrak{H}$ eineindeutig auf die obere Halbebene der komplexen z-Ebene ab.

Jedes solche Funktionensystem heiße ein Poincarésches Koordinatensystem auf $\mathfrak{H}$. Sind $z(p) = x_1(p) + i x_2(p)$, $z^*(p) = x_1^*(p) + i x_2^*(p)$ zwei Poincarésche Koordinatensysteme auf $\mathfrak{H}$, so gibt es solche reelle Konstanten a, b, c, d ($ad-bc = 1$), daß entweder

$$z^*(p) = \frac{az(p)+b}{cz(p)+d} \text{ oder } -\overline{z^*(p)} = \frac{az(p)+b}{cz(p)+d} .$$

---

[11]) Die Abschnitte 2.1 bis 2.11 enthalten, zur Bequemlichkeit des Lesers, eine kurze Zusammenstellung bekannter Tatsachen.

2.2. Ist $z(p) = x_1(p) + i x_2(p)$ ein Poincarésches Koordinatensystem, so führt die Abbildung $p \to z(p)$ die Gesamtheit der Geodätischen von $\mathfrak{H}$ über in die Gesamtheit der zur reellen Achse orthogonalen Halbkreise und Halbgeraden der oberen z-Halbebene.

2.3. Ist T eine (eigentliche) Bewegung von $\mathfrak{H}$ und $z(p) = x_1(p) + i x_2(p)$ irgendein Poincarésches Koordinatensystem auf $\mathfrak{H}$, so gibt es solche reelle Konstanten a, b, c, d $(ad - bc = 1)$, daß

$$z(Tp) = \frac{az(p) + b}{cz(p) + d}.$$

Die Zahl $|a + d|$ ist unabhängig vom Koordinatensystem $z(p)$, und es gilt

$|a + d| > 2$ genau dann, wenn T eine Translation $\neq E$ ist,
$|a + d| = 2$ genau dann, wenn T eine Grenzdrehung oder die Identität E ist,
$|a + d| < 2$ genau dann, wenn T eine Drehung $\neq E$ ist.

2.4. Für jedes Poincarésche Koordinatensystem $z(p) = x_1(p) + i x_2(p)$ auf $\mathfrak{H}$ gilt

$$\operatorname{Cos} \varrho(p, q) = 1 + \frac{(x_1(p) - x_1(q))^2 + (x_2(p) - x_2(q))^2}{2 x_2(p)\, x_2(q)}.$$

2.5. Aus 2.3 und 2.4 folgt: Für jedes feste $q \in \mathfrak{H}$ und jede Bewegung T ist $\operatorname{Cos} \varrho(Tp, q)$ eine auf $\mathfrak{H}$ reell-analytische Funktion von p.

2.6. Ist $T \neq E$ eine Translation von $\mathfrak{H}$, so gibt es auf $\mathfrak{H}$ ein solches Poincarésches Koordinatensystem $z(p) = x_1(p) + i x_2(p)$, daß

$$z(Tp) = \alpha z(p), \quad \alpha > 1. \tag{2}$$

Hieraus ergibt sich wegen 2.2 sofort, daß es auf $\mathfrak{H}$ eine einzige Geodätische [nämlich $x_1(p) = 0$] gibt, welche durch T in sich selbst übergeführt wird. Aus (2) folgt für jede ganze Zahl m: $z(T^m p) = \alpha^m z(p)$. Daraus und aus 2.4 ergibt sich leicht

$$\operatorname{Cos} \varrho(T^m p, p) = 1 + \left(\frac{\alpha^m + \alpha^{-m}}{2} - 1\right)\left(1 + \frac{x_1^2(p)}{x_2^2(p)}\right).$$

Hieraus erkennt man sofort

$$\operatorname{Cos} \mu(T^m) = \operatorname{Cos} \inf_{p \in \mathfrak{H}} \varrho(T^m p, p) = \inf_{p \in \mathfrak{H}} \operatorname{Cos} \varrho(T^m p, p) = \frac{\alpha^m + \alpha^{-m}}{2}.$$

Daher gilt wegen $\alpha > 1$: $\mu(T^m) = |m| \log \alpha$ und somit

$$\mu(T) = \log \alpha > 0 \tag{3}$$

$$\mu(T^m) = |m|\, \mu(T) \tag{4}$$

$$\operatorname{Cos} \varrho(T^m p, p) = 1 + (\operatorname{Cos} \mu(T^m) - 1)\left(1 + \frac{x_1^2(p)}{x_2^2(p)}\right). \tag{5}$$

2.7. Ist T eine Grenzdrehung von $\mathfrak{H}$, so gibt es auf $\mathfrak{H}$ ein solches Poincarésches Koordinatensystem $z(p) = x_1(p) + i x_2(p)$, daß entweder

$$z(Tp) = z(p) + 1 \quad \text{oder} \quad z(Tp) = z(p) - 1.$$

Dann ist nach 2.4

$$\operatorname{Cos} \varrho(Tp, p) = 1 + \frac{1}{2 x_2^2(p)}$$

und somit

$$\mu(T) = \inf_{p \in \mathfrak{H}} \varrho(Tp, p) = 0 .$$

2.8. Ist T eine Drehung von $\mathfrak{H}$, so ist trivialerweise $\mu(T) = 0$, da T auf $\mathfrak{H}$ einen Fixpunkt besitzt. Somit ergibt sich aus 2.6 und 2.7, daß die Translationen $T \neq E$ die einzigen Bewegungen von $\mathfrak{H}$ mit positiver Verschiebungslänge $\mu(T)$ sind.

2.9. Oft wird es zweckmäßig sein, statt Poincaréscher Koordinaten sog. geodätische Polarkoordinaten mit dem Pol q zu verwenden: Sei $\vartheta(p)$,

$$0 \leqq \vartheta(p) < 2\pi , \quad p \in \mathfrak{H} - q$$

der in positivem Sinne gemessene Winkel zwischen dem von q nach p führenden geodätischen Strahl und einem festen Strahl durch q. Dann ist

$$\{\varrho(p) = \varrho(p, q), \vartheta(p)\}$$

ein überall auf $\mathfrak{H} - q$ reguläres Koordinatensystem, bezüglich dessen die metrische Differentialform von $\mathfrak{H}$ die Gestalt

$$ds^2 = d\varrho^2 + \operatorname{Sin}^2 \varrho \, d\vartheta^2 \tag{6}$$

besitzt.

2.10. Es sei $K[q, r]$ die hyperbolische Kreisscheibe mit dem Zentrum q und dem Radius r, i. e. die Punktmenge $\{p \mid \varrho(p, q) \leqq r\}$. Dann ergibt sich aus (6)

$$\iint_{K[q,r]} d\omega = \int_{\vartheta=0}^{2\pi} \int_{\varrho=0}^{r} \operatorname{Sin} \varrho \, d\varrho \, d\vartheta = 2\pi(\operatorname{Cos} r - 1) .$$

2.11. Es sei $\{\xi(p), \eta(p)\}$ irgendein lokales reguläres Koordinatensystem auf $\mathfrak{H}$, und es sei $ds^2 = E(\xi, \eta)\, d\xi^2 + 2F(\xi, \eta)\, d\xi\, d\eta + G(\xi, \eta)\, d\eta^2$ die zugehörige metrische Differentialform von $\mathfrak{H}$. Dann drücken sich die beiden zu $\mathfrak{H}$ gehörigen Beltramischen Differentialoperatoren ∇_p und Δ_p im Koordinatensystem $\{\xi(p), \eta(p)\}$ folgendermaßen aus:

$$\nabla_p f(p) = \frac{E\left(\frac{\partial f}{\partial \eta}\right)^2 - 2F \frac{\partial f}{\partial \xi} \frac{\partial f}{\partial \eta} + G\left(\frac{\partial f}{\partial \xi}\right)^2}{EG - F^2}$$

$$\Delta_p f(p) = \frac{1}{\sqrt{EG - F^2}} \left\{ \frac{\partial}{\partial \eta} \left(\frac{E \frac{\partial f}{\partial \eta} - F \frac{\partial f}{\partial \xi}}{\sqrt{EG - F^2}} \right) + \frac{\partial}{\partial \xi} \left(\frac{G \frac{\partial f}{\partial \xi} - F \frac{\partial f}{\partial \eta}}{\sqrt{EG - F^2}} \right) \right\} .$$

2.12. Es sei T eine Bewegung von $\mathfrak{H}$ und $s = \sigma + it$. Dann gilt

(a) $\Delta_p \operatorname{Cos}^{-s} \varrho(Tp, q) = s(s-1) \operatorname{Cos}^{-s} \varrho(Tp, q) - s(s+1) \operatorname{Cos}^{-(s+2)} \varrho(Tp, q)$

(b) $\Delta_p \operatorname{Cos} \varrho(Tp, q) \quad = 2 \operatorname{Cos} \varrho(Tp, q)$

(c) $\nabla_p \operatorname{Cos} \varrho(Tp, q) \quad = \operatorname{Sin}^2 \varrho(Tp, q)$

(d) $\nabla_p^2 \operatorname{Cos} \varrho(Tp, q) \quad = 4 \operatorname{Sin}^2 \varrho(Tp, q) \operatorname{Cos}^2 \varrho(Tp, q)$.

2.13. Beweis: Wegen der Bewegungsinvarianz der Beltramischen Operatoren gilt

$$\begin{aligned} \Delta_p \operatorname{Cos}^{-s} \varrho(Tp, q) &= (\Delta_u \operatorname{Cos}^{-s} \varrho(u, q))_{u = Tp} \\ \nabla_p \operatorname{Cos} \varrho(Tp, q) &= (\nabla_u \operatorname{Cos} \varrho(u, q))_{u = Tp} \\ \nabla_p^2 \operatorname{Cos} \varrho(Tp, q) &= (\nabla_u^2 \operatorname{Cos} \varrho(u, q))_{u = Tp} . \end{aligned} \tag{7}$$

Zur weiteren Berechnung der rechten Seiten führen wir gemäß 2.9 auf $\mathfrak{H}$ geodätische Polarkoordinaten mit dem Pol q ein:

$$\varrho(u) = \varrho(u, q), \quad \vartheta(u), \quad u \in \mathfrak{H} - q. \tag{8}$$

Weil die metrische Differentialform von $\mathfrak{H}$ im Koordinatensystem $\{\varrho(u), \vartheta(u)\}$ die Gestalt $d\varrho^2 + \mathrm{Sin}^2\varrho\, d\vartheta^2$ besitzt, so drücken sich die Beltramischen Operatoren nach 2.11 in diesem System folgendermaßen aus:

$$\begin{aligned} \Delta_u &= \frac{1}{\mathrm{Sin}\,\varrho} \frac{\partial}{\partial\varrho}\left(\mathrm{Sin}\,\varrho \frac{\partial}{\partial\varrho}\right) + \frac{1}{\mathrm{Sin}^2\varrho} \frac{\partial^2}{\partial\vartheta^2}, \quad u \in \mathfrak{H} \\ \nabla_u &= \left(\frac{\partial}{\partial\varrho}\right)^2 + \frac{1}{\mathrm{Sin}^2\varrho}\left(\frac{\partial}{\partial\vartheta}\right)^2, \quad u \in \mathfrak{H}. \end{aligned} \tag{9}$$

Aus (7), (8) und (9) ergeben sich ohne weiteres die Gleichungen (a) bis (d) für $p \neq T^{-1}q$. Die linken und rechten Seiten dieser Gleichungen sind aber wegen 2.5 überall auf $\mathfrak{H}$ stetige Funktionen von p. Folglich müssen die Gleichungen (a) bis (d) auch noch für $p = T^{-1}q$ gelten. q.e.d.

2.14. Die reelle Funktion $f(p)$ sei auf $\mathfrak{H}$ zweimal stetig differenzierbar bezüglich der Poincaréschen Koordination $x_1(p)$, $x_2(p)$. Dann gilt[12])

$$\left|\frac{\partial f(p)}{\partial x_i}\right| \leqq x_2^{-1}(p)\sqrt{\nabla_p f(p)}$$

und für $\nabla_p f(p) \neq 0$:

$$\left|\frac{\partial^2 f(p)}{\partial x_i \partial x_k}\right| \leqq x_2^{-2}(p)\left(\sqrt{\frac{\nabla_p^2 f(p)}{\nabla_p f(p)}} + \sqrt{\nabla_p f(p)} + |\Delta_p f(p)|\right).$$

2.15. Aus 2.5, 2.12, 2.14 ergibt sich

$$\left|\frac{\partial}{\partial x_i} \mathrm{Cos}\,\varrho(Tp, q)\right| \leqq x_2^{-1}(p)\,\mathrm{Sin}\,\varrho(Tp, q) < x_2^{-1}(p)\,\mathrm{Cos}\,\varrho(Tp, q) \tag{10}$$

$$\begin{aligned} \left|\frac{\partial^2}{\partial x_i \partial x_k} \mathrm{Cos}\,\varrho(Tp, q)\right| &\leqq x_2^{-2}(p)\,(4\,\mathrm{Cos}\,\varrho(Tp, q) + \mathrm{Sin}\,\varrho(Tp, q) < \\ &< 5\,x_2^{-2}(p)\,\mathrm{Cos}\,\varrho(Tp, q). \end{aligned} \tag{11}$$

Nun ist

$$\frac{\partial}{x_i} \mathrm{Cos}^{-s}\varrho(Tp, q) = -s\,\mathrm{Cos}^{-s-1}\varrho(Tp, q) \frac{\partial}{\partial x_i} \mathrm{Cos}\,\varrho(Tp, q)$$

$$\begin{aligned} \frac{\partial^2}{\partial x_i \partial x_k} \mathrm{Cos}^{-s}\varrho(Tp, q) &= -s\,\mathrm{Cos}^{-s-1}\varrho(Tp, q) \frac{\partial}{\partial x_i \partial x_k} \mathrm{Cos}\,\varrho(Tp, q) + \\ &+ s(s+1)\,\mathrm{Cos}^{-s-2}\varrho(Tp, q) \frac{\partial}{\partial x_i} \mathrm{Cos}\,\varrho(Tp, q) \frac{\partial}{\partial x_k} \mathrm{Cos}\,\varrho(Tp, q). \end{aligned}$$

Daraus und aus (10), (11) ergibt sich sofort

2.16. Es sei T eine Bewegung von $\mathfrak{H}$, $s = \sigma + it$, und $\{x_1(p), x_2(p)\}$ ein Poincarésches Koordinatensystem auf $\mathfrak{H}$. Dann gilt:

$$\left|\frac{\partial}{\partial x_i} \mathrm{Cos}^{-s}\varrho(Tp, q)\right| < |s|\, x_2^{-1}(p)\,\mathrm{Cos}^{-\sigma}\varrho(Tp, q)$$

$$\left|\frac{\partial^2}{\partial x_i \partial x_k} \mathrm{Cos}^{-s}\varrho(Tp, q)\right| < |s|\,(5 + |s+1|)\, x_2^{-2}(p)\,\mathrm{Cos}^{-\sigma}\varrho(Tp, q).$$

---

12) [4], pag. 35, Lemma 2.

3. Beweis von Satz 1

3.1. Es sei $N(p, q, t)$ die Anzahl der Elemente der Menge

$$\Theta = \{T \mid T \in \Lambda, \varrho(Tp, q) \leqq t\}.$$

Es gibt eine nur von $\mathfrak{F}$ abhängige positive Konstante c_0 so, daß für alle $t \geqq 0$ $N(p, q, t) < c_0 \operatorname{Cos} t$.

3.2. Beweis: Da $\mathfrak{F}$ eine geschlossene Fläche ist, gilt offensichtlich $\mu_0 = \inf\limits_{\mathfrak{W} \neq 0} \mu(\mathfrak{W}) > 0$. Nach 1.2 ist somit $\varrho(Tp, p) \geqq \mu_0 > 0$ für alle $p \in \mathfrak{H}$ und $T \in \Lambda - E$. Daraus folgt aber leicht, daß die Kreisscheiben $K[Tp, \mu_0/2]$, $T \in \Lambda$, paarweise nicht überlappen. Für $T \in \Theta$ liegt aber die Kreisscheibe $K[Tp, \mu_0/2]$ in der Kreisscheibe $K[q, t + \mu_0/2]$. Somit gilt

$$\sum_{T \in \Theta} \iint\limits_{K[Tp, \mu_0/2]} d\omega \leqq \iint\limits_{K[q, t+\mu_0/2]} d\omega,$$

also nach 2.10: $2\pi(\operatorname{Cos}\mu_0/2 - 1)\, N(p, q, t) \leqq 2\pi(\operatorname{Cos}(t + \mu_0/2) - 1)$. Daraus folgt

$$N(p, q, t) \leqq \frac{\operatorname{Cos}(t + \mu_0/2) - 1}{\operatorname{Cos}\mu_0/2 - 1} < \frac{\operatorname{Cos}(t + \mu_0/2)}{\operatorname{Cos}\mu_0/2 - 1} < \frac{e^{\mu_0/2}}{\operatorname{Cos}\mu_0/2 - 1} \operatorname{Cos} t$$

q.e.d.

3.3. Für $\sigma > 1$ und $0 \leqq t \leqq t_1$ gilt

$$\sum_{t \leqq \varrho(Tp, q) \leqq t_1} \operatorname{Cos}^{-\sigma} \varrho(Tp, q) < c_0\left(1 + \frac{1}{\sigma - 1}\right) \operatorname{Cos}^{-(\sigma-1)} t.$$

3.4. Beweis: Offenbar gilt

$$\sum_{t \leqq \varrho(Tp, q) \leqq t_1} \operatorname{Cos}^{-\sigma} \varrho(Tp, q) = N(p, q, t) \operatorname{Cos}^{-\sigma} t + \int_t^{t_1} \operatorname{Cos}^{-\sigma} \tau \, dN(p, q, \tau).$$

Partielle Integration ergibt

$$\sum_{t \leqq \varrho(Tp, q) \leqq t_1} \operatorname{Cos}^{-\sigma} \varrho(Tp, q) = N(p, q, t_1) \operatorname{Cos}^{-\sigma} t_1 + \sigma \int_t^{t_1} N(p, q, \tau) \operatorname{Cos}^{-\sigma-1} \tau \operatorname{Sin} \tau \, d\tau.$$

Daraus und aus 3.1 folgt für $\sigma > 1$

$$\sum_{t \leqq \varrho(Tp, q) \leqq t_1} \operatorname{Cos}^{-\sigma} \varrho(Tp, q) < c_0 \operatorname{Cos}^{-(\sigma-1)} t_1 + c_0 \sigma \int_t^{t_1} \operatorname{Cos}^{-\sigma} \tau \operatorname{Sin} \tau \, d\tau$$

$$= -\frac{c_0}{\sigma - 1} \operatorname{Cos}^{-(\sigma-1)} t_1 + c_0\left(1 + \frac{1}{\sigma-1}\right) \operatorname{Cos}^{-(\sigma-1)} t < c_0\left(1 + \frac{1}{\sigma - 1}\right) \operatorname{Cos}^{-(\sigma-1)} t.$$

3.5. Es sei $\{T_n \mid n \geqq 1\}$ eine beliebige Anordnung der Elemente der Gruppe Λ und $\sigma > 1$. Dann konvergiert die Reihe

$$\sum_{n=1}^{\infty} \operatorname{Cos}^{-\sigma} \varrho(T_n p, q)$$

gleichmäßig in (p, q) auf jedem Kompaktum von $\mathfrak{H}$.

3.6. Beweis: Sei $p_0 \in \mathfrak{H}$ und $r > 0$. Die Behauptung wird bewiesen sein, wenn wir die gleichmäßige Konvergenz für alle $p, q \in K[p_0, r]$ zeigen können.

Wegen $\sigma > 1$ gibt es zu jedem $\varepsilon > 0$ ein solches $t_\varepsilon > 0$, daß

(1) $$c_0\left(1 + \frac{1}{\sigma - 1}\right) \mathrm{Cos}^{-(\sigma-1)}\, t_\varepsilon < \varepsilon .$$

Wegen der Diskontinuität von Λ gibt es einen solchen Index $j_0 = j_0(\varepsilon, r)$, daß $T_j(K[p_0, r]) \cap K[p_0, r + t_\varepsilon] = 0$ für alle $j > j_0(\varepsilon, r)$. Dann gilt für $p, q \in K[p_0, r]$ und $j > j_0(\varepsilon, r)$:

$$\begin{aligned} 2r + t_\varepsilon < \varrho(T_j p_0, p_0) &\leqq \varrho(T_j p_0, T_j p) + \varrho(T_j p, q) + \varrho(q, p_0) \\ &= \varrho(T_j p, q) + \varrho(p_0, p) + \varrho(p_0, q) \leqq \varrho(T_j p, q) + 2r , \end{aligned}$$

also $\varrho(T_j p, q) > t_\varepsilon$. Somit gilt nach 3.3 und (1)

$$\sum_{j=m}^{n} \mathrm{Cos}^{-\sigma} \varrho(T_j p, q) < c_0\left(1 + \frac{1}{\sigma - 1}\right) \mathrm{Cos}^{-(\sigma-1)}\, t_\varepsilon < \varepsilon$$

für alle $p, q \in K[p_0, r]$ und $n \geqq m > j_0(\varepsilon, r)$. q.e.d.

3.7. Aus 3.6 und 2.16 folgt: Es sei $\{x_1(p), x_2(p)\}$ ein Poincarésches Koordinatensystem auf $\mathfrak{H}$, $\Lambda = \{T_n \mid n \geqq 1\}$, $s = \sigma + it$, $\sigma > 1$. Dann konvergieren die Reihen

$$\sum_{n=1}^{\infty} \mathrm{Cos}^{-s} \varrho(T_n p, q)\,, \quad \sum_{n=1}^{\infty} \frac{\partial}{\partial x_i} \mathrm{Cos}^{-s} \varrho(T_n p, q)\,, \quad \sum_{n=1}^{\infty} \frac{\partial^2}{\partial x_i \partial x_k} \mathrm{Cos}^{-s} \varrho(T_n p, q)$$

absolut und gleichmäßig in (p, q) auf jedem Kompaktum von $\mathfrak{H}$.

3.8. Aus 3.7 und 2.5 folgt: Für alle $s = \sigma + it$, $\sigma > 1$, und $q \in \mathfrak{H}$ gilt:

(a) $G(p, q; s) = \sum_{T \in \Lambda} \mathrm{Cos}^{-s} \varrho(Tp, q)$ ist eine auf $\mathfrak{H}$ stetige und bezüglich Λ automorphe Funktion von p.

(b) $G(p, q; s)$ ist überall auf $\mathfrak{H}$ zweimal stetig differenzierbar nach Poincaréschen Koordinaten $x_1(p)$, $x_2(p)$, und es gilt

$$\frac{\partial}{\partial x_i} G(p, q; s) = \sum_{T \in \Lambda} \frac{\partial}{\partial x_i} \mathrm{Cos}^{-s} \varrho(Tp, q)\,, \quad \frac{\partial^2}{\partial x_i \partial x_k} G(p, q; s) = \sum_{T \in \Lambda} \frac{\partial^2}{\partial x_i \partial x_k} \mathrm{Cos}^{-s} \varrho(Tp, q)\,.$$

3.9. Für $\sigma > 1$ gilt

$$\Delta_p G(p, q; s) + s(1 - s)\, G(p, q; s) + s(1 + s)\, G(p, q, s + 2) = 0\,.$$

3.10. Beweis: Sei $\{x_1(p), x_2(p)\}$ ein Poincarésches Koordinatensystem auf $\mathfrak{H}$. Dann ist nach 2.1 und 2.11

$$\Delta_p = x_2^2(p)\left(\frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2}\right).$$

Somit folgt aus 3.8 (b)

$$\Delta_p G(p, q, s) = \sum_{T \in \Lambda} \Delta_p \mathrm{Cos}^{-s} \varrho(Tp, q)\,, \quad \sigma > 1\,.$$

Hieraus und aus 2.12 (a) ergibt sich sofort unsere Behauptung.

4. Beweis von Satz 2 und Satz 4

4.1. Da die Fläche $\mathfrak{F}$ geschlossen ist, besitzt die Gruppe Λ kompakte Fundamentalbereiche; es gibt sogar einen kompakten Fundamentalbereich

$\mathfrak{B} \subset \mathfrak{H}$, welcher ein von endlich vielen geodätischen Strecken berandetes konvexes Polygon ist[13]).

Wir untersuchen jetzt die Fourierkoeffizienten der automorphen Funktion $G(p, q; s)$ bezüglich des in 1.6 eingeführten Orthonormalsystems $\{\varphi_n(p)\}$:

$$(1) \qquad a_n(s, q) = \iint_{\mathfrak{B}} G(p, q; s)\, \varphi_n(p)\, d\omega_p\,, \quad n \geqq 0\,, \quad \sigma > 1\,.$$

Da das Fundamentalpolygon $\mathfrak{B}$ kompakt ist, folgt aus 3.7

$$\begin{aligned} a_n(s, q) &= \sum_{T \in \Lambda} \iint_{\mathfrak{B}} \operatorname{Cos}^{-s} \varrho(Tp, q)\, \varphi_n(p)\, d\omega_p \\ (2) \qquad &= \sum_{T \in \Lambda} \iint_{\mathfrak{B}} \operatorname{Cos}^{-s} \varrho(Tp, q)\, \varphi_n(Tp)\, d\omega_p \\ &= \sum_{T \in \Lambda} \iint_{T(\mathfrak{B})} \operatorname{Cos}^{-s} \varrho(p, q)\, \varphi_n(p)\, d\omega_p\,, \quad \sigma > 1\,. \end{aligned}$$

Dabei wurde benutzt, daß $\varphi_n(p)$ automorph bezüglich Λ ist, und daß das Flächenelement $d\omega$ invariant gegenüber den Bewegungen T ist. In derselben Weise ergibt sich

$$\begin{aligned} \iint_{\mathfrak{B}} G(p, q, \sigma)\, |\varphi_n(p)|\, d\omega_p &= \sum_{T \in \Lambda} \iint_{T(\mathfrak{B})} \operatorname{Cos}^{-\sigma} \varrho(p, q)\, |\varphi_n(p)|\, d\omega_p \\ &= \sum_{T \in \Lambda} \iint_{T(\mathfrak{B})} |\operatorname{Cos}^{-s} \varrho(p, q)\, \varphi_n(p)|\, d\omega_p\,, \quad \sigma > 1\,, \end{aligned}$$

also insbesondere

$$\sum_{T \in \Lambda} \iint_{T(\mathfrak{B})} |\operatorname{Cos}^{-s} \varrho(p, q)\, \varphi_n(p)|\, d\omega_p < +\infty \quad \text{für} \quad \sigma > 1\,.$$

Daraus folgt, da die Fundamentalpolygone $T(\mathfrak{B})$, $T \in \Lambda$, ganz $\mathfrak{H}$ überdecken und paarweise nicht überlappen, daß $\operatorname{Cos}^{-s} \varrho(p, q)\, \varphi_n(p)$ für $\sigma > 1$ über $\mathfrak{H}$ integrierbar ist und daß gilt:

$$\sum_{T \in \Lambda} \iint_{T(\mathfrak{B})} \operatorname{Cos}^{-s} \varrho(p, q)\, \varphi_n(p)\, d\omega_p = \iint_{\mathfrak{H}} \operatorname{Cos}^{-s} \varrho(p, q)\, \varphi_n(p)\, d\omega\,.$$

Somit folgt aus (2)

$$(3) \qquad a_n(s, q) = \iint_{\mathfrak{H}} \operatorname{Cos}^{-s} \varrho(p, q)\, \varphi_n(p)\, d\omega\,, \quad \sigma > 1\,.$$

Zur weiteren Berechnung dieses Integrals führen wir auf $\mathfrak{H} - q$ in folgender Weise Koordinaten ein: Es sei $\vartheta(p)$,

$$0 \leqq \vartheta(p) < 2\pi\,, \quad p \in \mathfrak{H} - q\,,$$

der in positivem Sinne gemessene Winkel zwischen dem geodätischen Strahl von q nach p und einem festen Strahl durch q, und

$$(4) \qquad \tau(p) = \log \operatorname{Cos} \varrho(p, q)\,.$$

Wie man auf Grund von 2.9 leicht nachrechnet, drückt sich die metrische Differentialform von $\mathfrak{H}$ in dem auf $\mathfrak{H} - q$ überall regulären Koordinatensystem $\{\tau(p), \vartheta(p)\}$ folgendermaßen aus:

$$ds^2 = \frac{e^{2\tau}}{e^{2\tau} - 1}\, d\tau^2 + (e^{2\tau} - 1)\, d\vartheta^2\,;$$

---

[13]) Siehe Anmerkung [1]).

daher ist

(5) $$d\omega_p = e^\tau d\tau\, d\vartheta\,.$$

Nun definieren wir

(6) $$\begin{aligned} \psi_n(\tau,\vartheta) &= \varphi_n(p(\tau,\vartheta)) \quad \text{für} \quad 0<\tau<\infty,\ 0\leqq\vartheta<2\pi \\ \psi_n(0,\vartheta) &= \varphi_n(q) \quad \text{für} \quad 0\leqq\vartheta<2\pi \end{aligned}$$

(7) $$\Phi_n(\tau) = \int_0^{2\pi} \psi_n(\tau,\vartheta)\, d\vartheta\,.$$

Da $\varphi_n(p)$ auf $\mathfrak{H}$ stetig und beschränkt ist, gilt offenbar:

4.2. $\Phi_n(\tau)$ ist stetig und beschränkt für $\tau\geqq 0$ und es gilt

$$\Phi_n(0) = 2\pi\, \varphi_n(q)\,.$$

4.3. Aus (3) bis (7) ergibt sich jetzt

$$a_n(s,q) = \int_0^\infty e^{-(s-1)\tau}\, \Phi_n(\tau)\, d\tau\,.$$

Daraus und aus 4.2 folgt nach bekannten Sätzen[14]) aus der Theorie der Laplace-Transformation:

4.4. $a_n(s,q)$ ist regulär analytisch in der Halbebene $\sigma>1$, und es gilt

$$\lim_{\sigma\to+\infty} \sigma a_n(\sigma,q) = 2\pi\, \varphi_n(q)\,.$$

4.5. Die Funktion

(8) $$f_n(s) = a_n(s,q)\, 2^{-s}\, \Gamma(s) \Big/ \Gamma\left(\frac{s-s^+(\lambda_n')}{2}\right) \Gamma\left(\frac{s-s^-(\lambda_n')}{2}\right)$$

ist regulär analytisch in der Halbebene $\sigma>1$, und es gilt dort

$$f_n(s+2) = f_n(s)\,.$$

4.6. Beweis: Da $1/\Gamma(s)$ eine ganze Funktion ist, die nur in den Punkten $s=-l,\ l\geqq 0$ ganz, verschwindet, so folgt aus (8) und 4.4 in der Tat, daß $f_n(s)$ regulär in der Halbebene $\sigma>1$ ist.

Wegen (1) und 3.9 gilt für $\sigma>1$

(9) $$\iint_{\mathfrak{B}} \varphi_n(p)\, \Delta_p\, G(p,q;s)\, d\omega_p + s(1-s)\, a_n(s,q) + s(1+s)\, a_n(s+2,q) = 0\,.$$

Weil die Fläche $\mathfrak{F}$ geschlossen ist, und weil $\varphi_n(p)$ und $G(p,q;s)$ zweimal stetig differenzierbare und bezüglich Λ automorphe Funktionen von p sind, so ergibt die Greensche Formel

$$\begin{aligned} \iint_{\mathfrak{B}} \varphi_n(p)\, \Delta_p G(p,q;s)\, d\omega_n &= \iint_{\mathfrak{B}} G(p,q;s)\, \Delta_p\, \varphi_n(p)\, d\omega_p \\ &= -\lambda_n' \iint_{\mathfrak{B}} G(p,q;s)\, \varphi_n(p)\, d\omega = -\lambda_n' a_n(s,q)\,. \end{aligned}$$

Daraus und aus (9) folgt: $s(s+1)\, a_n(s+2,q) = (s^2 - s + \lambda_n')\, a_n(s,q)$; somit gilt wegen 1.4 (8)

$$s(s+1)\, a_n(s+2,q) = (s - s^+(\lambda_n'))\, (s - s^-(\lambda_n'))\, a_n(s,q)\,.$$

[14]) [2], pag. 473, Satz 1.

Hieraus und aus (8) folgt nun unter Berücksichtigung von $\Gamma(s+1) = s\,\Gamma(s)$ in der Tat: $f_n(s+2) = f_n(s)$.

4.7. Aus 1.4 (8) und aus der Stirlingschen asymptotischen Darstellung von $\Gamma(s)$ ergibt sich leicht

$$2^{-\sigma}\Gamma(\sigma)\Big/\Gamma\left(\frac{\sigma-s^+(\lambda_n')}{2}\right)\Gamma\left(\frac{\sigma-s^-(\lambda_n')}{2}\right) \sim \frac{1}{4\sqrt{\pi}}\,\sigma\,, \quad \sigma\to+\infty\,.$$

Somit folgt aus (8) und 4.4

$$\lim_{\sigma\to+\infty} f_n(\sigma) = \frac{\sqrt{\pi}}{2}\,\varphi_n(q)\,.$$

Daraus und aus 4.5 folgt nun offensichtlich $f_n(s) = \text{const} = \frac{\sqrt{\pi}}{2}\,\varphi_n(q)$. Damit haben wir bewiesen:

4.8. Es sei

$$a_n(s,q) = \iint_{\mathfrak{B}} G(p,q;s)\,\varphi_n(p)\,d\omega_p\,, \quad \sigma>1\,, \quad n\geqq 0$$

$$F(s,\lambda) = \Gamma\left(\frac{s-s^+(\lambda)}{2}\right)\Gamma\left(\frac{s-s^-(\lambda)}{2}\right),\quad \lambda\geqq 0\,. \tag{10}$$

Dann gilt für $\sigma>1$

$$a_n(s,q) = \sqrt{\pi}\,\frac{2^{s-1}}{\Gamma(s)}\,F(s,\lambda_n')\,\varphi_n(q)\,.$$

4.9. Es gibt eine nur von $\mathfrak{F}$ abhängige Konstante $c_1>0$ derart, daß

$$\sum_{n=1}^{\infty}|F(s,\lambda_n')\,\varphi_n(q)|^2 \leqq c_1\Gamma^2(\sigma_0) \quad \text{für} \quad 2\leqq\sigma\leqq\sigma_0\,.$$

4.10. Beweis: Da $\{\varphi_n(p)\}$ ein Orthonormalsystem ist, gilt die Besselsche Ungleichung

$$\sum_{n=1}^{\infty}|a_n(s,q)|^2 \leqq \iint_{\mathfrak{B}} |G(p,q;s)|^2\,d\omega_p \quad \text{für} \quad \sigma>1\,.$$

Daraus folgt nach 4.8

$$\begin{aligned}\sum_{n=1}^{\infty}|F(s,\lambda_n')\,\varphi_n(q)|^2 &\leqq \frac{|\Gamma(s)|^2}{\pi\,2^{2\sigma-2}}\iint_{\mathfrak{B}}|G(p,q;s)|^2\,d\omega_p \leqq \\ &\leqq \frac{|\Gamma(s)|^2}{4\pi}\iint_{\mathfrak{B}}|G(p,q;s)|^2\,d\omega_p\,, \quad \sigma\geqq 2\,.\end{aligned} \tag{11}$$

Aus der für $\sigma>0$ gültigen Darstellung $\Gamma(s) = \int\limits_0^\infty e^{-t}t^{s-1}\,dt$ ist zu entnehmen, daß $|\Gamma(s)|\leqq\Gamma(\sigma)$ für $\sigma>0$. Da außerdem $\Gamma(\sigma)$ für $\sigma\geqq 2$ monoton wächst, so folgt aus (11)

$$\sum_{n=1}^{\infty}|F(s,\lambda_n')\,\varphi_n(q)|^2 \leqq \frac{1}{4\pi}\,\Gamma^2(\sigma_0)\iint_{\mathfrak{B}}|G(p,q,s)|^2\,d\omega_p \quad \text{für} \quad 2\leqq\sigma\leqq\sigma_0\,. \tag{12}$$

Nach 3.3 gilt

$$|G(p,q,s)| \leqq G(p,q;\sigma) \leqq c_0\left(1+\frac{1}{\sigma-1}\right) \quad \text{für} \quad \sigma>1\,.$$

Daher ist wegen 1.5 (10)

$$\iint\limits_{\mathfrak{G}} |G(p,q;s)|^2 d\omega_p \leq c_0^2\left(1+\frac{1}{\sigma-1}\right)^2 4\pi(g_{\mathfrak{F}}-1) \leq 16\pi\, c_0^2(g_{\mathfrak{F}}-1) \quad \text{für} \quad \sigma \geq 2\,.$$

Somit folgt aus (12)

$$\sum_{n=1}^{\infty} |F(s,\lambda_n')\, \varphi_n(q)|^2 \leq 4c_0^2(g_{\mathfrak{F}}-1)\,\Gamma^2(\sigma_0) \quad \text{für} \quad 2 \leq \sigma \leq \sigma_0\,.$$

Da die Konstante c_0 nach 3.1 nur von $\mathfrak{F}$ abhängt, ist damit unsere Behauptung bewiesen.

4.11. Es sei $h \geq 2$ ganz, $Q_h = \{s \mid |\sigma| \leq h,\ |t| \leq h\}$. Dann gilt für alle $s \in Q_h$ und $\lambda \geq 4h^2 + 1/4$

$$|F(s,\lambda)| \leq 80\left(\frac{4}{h^2}\right)^{h-1} \frac{1}{\lambda}\, |F(s+2h,\lambda)|\,.$$

4.12. Beweis: Wegen $\Gamma(s+1) = s\,\Gamma(s)$ erfüllt die Funktion (10) offenbar die Differenzengleichung

$$F(s,\lambda) = \frac{4}{(s-s^+(\lambda))(s-s^-(\lambda))}\, F(s+2,\lambda)\,.$$

Daraus ergibt sich für $h \geq 2$

$$F(s,\lambda) = \frac{4^h}{(s-s^+(\lambda))(s-s^-(\lambda))}\left(\prod_{k=1}^{h-1}(s+2k-s^+(\lambda))(s+2k-s^-(\lambda))\right)^{-1} \times\, F(s+2h,\lambda)\,.$$

Berücksichtigt man noch, daß nach 1.4 (8) $s^+(\lambda)\, s^-(\lambda) = \lambda$, so kommt

$$F(s,\lambda) = \frac{4^h\lambda^{-1}}{\left(1-\frac{s}{s^+(\lambda)}\right)\left(1-\frac{s}{s^-(\lambda)}\right)}\left(\prod_{k=1}^{h-1}(s+2k-s^+(\lambda))(s+2k-s^-(\lambda))\right)^{-1} \times\, F(s+2h,\lambda)\,. \tag{12}$$

Nach 1.4 (8) gilt

$$|\operatorname{Im} s^+(\lambda)| = |\operatorname{Im} s^-(\lambda)| \geq 2h \quad \text{für} \quad \lambda \geq 4h^2 + 1/4\,. \tag{13}$$

Daher gilt für $\lambda \geq 4h^2 + 1/4$, $s \in Q_h$, k ganz:

$$|s+2k-s^{\pm}(\lambda)| \geq |\operatorname{Im}(s+2k-s^{\pm}(\lambda))| = |\sigma - \operatorname{Im} s^{\pm}(\lambda)| \geq$$
$$\geq |\operatorname{Im} s^{\pm}(\lambda)| - |\sigma| = 2h - |\sigma| \geq h\,,$$

somit

$$\left|\prod_{k=1}^{h-1}(s+2k-s^+(\lambda))(s+2k-s^-(\lambda))\right| \geq h^{2h-2}, \quad s \in Q_h, \quad \lambda \geq 4h^2 + 1/4\,. \tag{14}$$

Wegen (13) gilt für $s \in Q_h$, $\lambda \geq 4h^2 + 1/4$ ferner

$$\left|\frac{s}{s^{\pm}(\lambda)}\right| \leq \frac{\sqrt{2}\,h}{|\operatorname{Im} s^{\pm}(\lambda)|} \leq \frac{\sqrt{2}}{2}$$

$$\left|1-\frac{s}{s^{\pm}(\lambda)}\right| \geq 1 - \left|\frac{s}{s^{\pm}(\lambda)}\right| \geq 1 - \frac{\sqrt{2}}{2} > 0\,,$$

somit

$$\left|\left(1-\frac{s}{s^+(\lambda)}\right)\left(1-\frac{s}{s^-(\lambda)}\right)\right| \geqq \left(1-\frac{\sqrt{2}}{2}\right)^2 > \frac{1}{20}\,, \quad s\in Q_h\,, \quad \lambda \geqq 4h^2+1/4. \tag{15}$$

Aus (12) und (14), (15) folgt aber die Behauptung.

4.13. Es sei $h \geqq 2$ ganz, $n_0(h) = \operatorname{Max}\{n \mid \lambda_n' \leqq 4h^2+1/4\}$ Dann gilt:

(a) Für alle $l \geqq k \geqq n_0(h)$, $s \in Q_h$, $p, q \in \mathfrak{H}$

$$\sum_{n=k}^{l} |F(s,\lambda_n')\,\varphi_n(p)\,\varphi_n(q)| \leqq c_2\left(\frac{4}{h^2}\right)^{h-1}\Gamma(3h)\left[\sum_{n=k}^{l}\left(\frac{\varphi_n(p)}{\lambda_n'}\right)^2\right]^{1/2}.$$

(b) Für alle $l \geqq k \geqq n_0(h)$, $s \in Q_h$

$$\sum_{n=k}^{l} |F(s,\lambda_n')| \leqq c_3\left(\frac{4}{h^2}\right)^{h-1}\Gamma(3h)\left[\sum_{n=k}^{l} 1/\lambda_n'^2\right]^{1/2}.$$

Dabei sind die Konstanten c_2, c_3 nur von $\mathfrak{F}$ abhängig.

4.14. Beweis: Für $l \geqq k \geqq n_0(h)$ und $s \in Q_h$ folgt nach 4.11

$$\sum_{n=k}^{l} |F(s,\lambda_n')\,\varphi_n(p)\,\varphi_n(q)| \leqq 80\left(\frac{4}{h^2}\right)^{h-1}\sum_{n=k}^{l}|F(s+2h,\lambda_n')\,\varphi_n(q)|\,\frac{|\varphi_n(p)|}{\lambda_n'} \leqq$$

$$\leqq 80\left(\frac{4}{h^2}\right)^{h-1}\left[\sum_{n=k}^{l}|F(s+2h,\lambda_n')\,\varphi_n(q)|^2\right]^{\frac12}\left[\sum_{n=k}^{l}\left(\frac{\varphi_n(p)}{\lambda_n'}\right)^2\right]^{\frac12}. \tag{16}$$

Für $s \in Q_h$, $h \geqq 2$, gilt aber $2 \leqq h \leqq Re(s+2h) \leqq 3h$. Daher ist nach 4.9

$$\sum_{n=k}^{l}|F(s+2h,\lambda_n')\,\varphi_n(q)|^2 \leqq c_1\Gamma^2(3h),$$

wobei die Konstante c_1 nur von $\mathfrak{F}$ abhängt. Somit folgt aus (16)

$$\sum_{n=k}^{l}|F(s,\lambda_n')\,\varphi_n(p)\,\varphi_n(q)| \leqq 80\sqrt{c_1}\left(\frac{4}{h^2}\right)^{h-1}\Gamma(3h)\left[\sum_{n=k}^{l}\left(\frac{\varphi_n(p)}{\lambda_n'}\right)^2\right]^{\frac12},$$

$$l \geqq k \geqq n_0(h), \quad s \in Q_h. \tag{17}$$

Damit ist die Behauptung (a) bewiesen. Aus (17) ergibt sich für $p = q$

$$\sum_{n=k}^{l}|F(s,\lambda_n')|\,|\varphi_n(p)|^2 \leqq 80\sqrt{c_1}\left(\frac{4}{h^2}\right)^{h-1}\Gamma(3h)\left[\sum_{n=k}^{l}\left(\frac{\varphi_n(p)}{\lambda_n'}\right)^2\right]^{\frac12},$$

$$l \geqq k \geqq n_0(h), \quad s \in Q_h.$$

Daraus folgt wegen $\iint_{\mathfrak{B}} |\varphi_n(p)|^2 d\omega = 1$:

$$\sum_{n=k}^{l}|F(s,\lambda_n')| \leqq 80\sqrt{c_1}\left(\frac{4}{h^2}\right)^{h-1}\Gamma(3h)\iint_{\mathfrak{B}}\left[\sum_{n=k}^{l}\left(\frac{\varphi_n(p)}{\lambda_n'}\right)^2\right]^{\frac12}d\omega \leqq$$

$$\leqq 80\sqrt{c_1}\left(\frac{4}{h^2}\right)^{h-1}\Gamma(3h)\left[\iint_{\mathfrak{B}}\sum_{n=k}^{l}\left(\frac{\varphi_n(p)}{\lambda_n'}\right)^2 d\omega\right]^{\frac12}\left[\iint_{\mathfrak{B}} d\omega\right]^{\frac12}$$

$$= 80\sqrt{4\pi(g_{\mathfrak{F}}-1)\,c_1}\left(\frac{4}{h^2}\right)^{h-1}\Gamma(3h)\left[\sum_{n=k}^{l}\left(\frac{1}{\lambda_n'}\right)^2\right]^{\frac12}.$$

Damit ist auch die Behauptung (b) bewiesen.

4.15. Aus 4.13 und 1.7 (a) folgt: Es sei $h \geq 2$ ganz, $Q_h = \{s \mid |\sigma| \leq h, |t| \leq h\}$, $n_0(h) = \text{Max}\,\{n \mid \lambda_n' \leq 4h^2 + 1/4\}$. Dann gilt

(a) Die Reihe

$$\sum_{n=n_0(h)}^{\infty} F(s, \lambda_n')\, \varphi_n(p)\, \varphi_n(q)$$

konvergiert absolut und gleichmäßig für alle $(p, q, s) \in \mathfrak{H} \times \mathfrak{H} \times Q_h$.

(b) Die Reihe

$$\sum_{n=n_0(h)}^{\infty} F(s, \lambda_n')$$

konvergiert absolut und gleichmäßig in Q_h.

4.16. Wie aus (10) sofort ersichtlich ist, liegen die Pole der meromorphen Funktion $F(s, \lambda)$, $\lambda \geq 0$, in den Punkten

$$s^+(\lambda) - 2l\,, \quad s^-(\lambda) - 2l\,, \qquad (l \geq 0 \text{ ganz})\,.$$

Daher ergeben sich aus 4.15 unter Berücksichtigung der in 1.4 eingeführten Bezeichnungen sofort die folgenden Sätze 4.17 und 4.18:

4.17. Die Reihe

$$H(p, q; s) = \sum_{n=1}^{\infty} F(s, \lambda_n')\, \varphi_n(p)\, \varphi_n(q) \tag{18}$$

konvergiert absolut für $s \notin P = \{s^+(\lambda_n) - 2l,\, s^-(\lambda_n) - 2l \mid n \geq 1, l \geq 0\}$ und stellt eine in der ganzen s-Ebene meromorphe Funktion dar, deren Pole in der Punktmenge P enthalten sind. $H(p, q; s)$ besitzt im Punkte $s^{\pm}(\lambda_n) - 2l \in P$ denselben Hauptteil wie die Funktion

$$F(s, \lambda_n) \sum_{\lambda_k' = \lambda_n} \varphi_k(p)\, \varphi_k(q)\,.$$

4.18. Die Reihe

$$H(s) = \sum_{n=1}^{\infty} F(s, \lambda_n') = \sum_{n=1}^{\infty} m_n F(s, \lambda_n) \tag{19}$$

konvergiert absolut für $s \notin P$ und stellt eine in der ganzen s-Ebene meromorphe Funktion mit der Polmenge P dar. $H(s)$ besitzt im Punkte $s^{\pm}(\lambda_n) - 2l \in P$ denselben Hauptteil wie die Funktion $m_n F(s, \lambda_n)$.

4.19. Weil $\iint_{\mathfrak{F}} \varphi_n^2(p)\, d\omega = 1$, so folgt aus (18), (19) wegen 4.15 (a):

$$\iint_{\mathfrak{F}} H(p, p, s)\, d\omega = H(s) \quad \text{für} \quad s \notin P\,.$$

4.20. Aus 3.8 und 4.8 folgt nach 1.7 (b) für $\sigma > 1$:

$$G(p, q, s) = \sum_{n=0}^{\infty} \sqrt{\pi}\, \frac{2^{s-1}}{\Gamma(s)}\, F(s, \lambda_n')\, \varphi_n(q)\, \varphi_n(p)\,.$$

Somit gilt wegen (18) und 1.6 (13)

$$G(p, q, s) = \frac{1}{4\sqrt{\pi}\,(g_{\mathfrak{F}} - 1)}\, \frac{2^{s-1}}{\Gamma(s)}\, F(s, \lambda_0') + \sqrt{\pi}\, \frac{2^{s-1}}{\Gamma(s)}\, H(p, q; s)\,, \quad \sigma > 1\,. \tag{20}$$

Nach (10) und 1.4 (9) ist aber

$$(21) \qquad F(s, \lambda_0') = \Gamma\left(\frac{s}{2}\right)\Gamma\left(\frac{s-1}{2}\right) = \frac{2}{s-1}\Gamma\left(\frac{s}{2}\right)\Gamma\left(\frac{s+1}{2}\right).$$

Ferner gilt die Formel von Gauss-Legendre

$$(22) \qquad \sqrt{\pi}\,\Gamma(s) = 2^{s-1}\,\Gamma\left(\frac{s}{2}\right)\Gamma\left(\frac{s+1}{2}\right).$$

Aus (20), (21), (22) ergibt sich nun

4.21. In der Halbebene $\sigma > 1$ gilt

$$G(p, q; s) = \frac{1}{2(g_{\mathfrak{F}} - 1)}\,\frac{1}{s-1} + \sqrt{\pi}\,\frac{2^{s-1}}{\Gamma(s)}\,H(p, q, s).$$

4.22. Aus 4.19 und 4.21 folgt: In der Halbebene $\sigma > 1$ gilt

$$\iint_{\mathfrak{B}} G(p, p; s)\, d\omega_p = \frac{2\pi}{s-1} + \sqrt{\pi}\,\frac{2^{s-1}}{\Gamma(s)}\,H(s).$$

5. Beweis von Satz 5

5.1. Nach 2.6 führt jede Translation $T \neq E$ genau eine Geodätische auf $\mathfrak{H}$ in sich selbst über; diese invariante Geodätische a_T heiße die Achse von T. Offenbar gilt

$$(1) \qquad a_{T^m} = a_T \quad \text{für} \quad m \neq 0$$

$$(2) \qquad a_{STS^{-1}} = S(a_T).$$

Eine Geodätische a heiße Achse der Gruppe Λ, wenn $a = a_T$ für mindestens ein $T \in \Lambda - E$. Aus der Diskontinuität von Λ folgt leicht[15]):

5.2. Es sei a eine Achse von Λ. Dann ist $\Lambda\{a\} = \{T \mid T \in \Lambda, T(a) = a\}$ eine zyklische Untergruppe unendlicher Ordnung von Λ.

5.3. Ein Element $P \in \Lambda$ heiße Primelement, wenn es keine Darstellung $P = Q^n$ mit $n > 1$ und $Q \in \Lambda$ besitzt. Das Inverse eines Primelementes und die zu einem Primelement konjugierten Elemente von Λ sind offenbar ebenfalls Primelemente; eine aus Primelementen bestehende Klasse konjugierter Elemente von Λ möge Primklasse heißen.

Aus 5.2 und (1) folgt sofort: Ein Element $P \in \Lambda - E$ ist genau dann Primelement, wenn es eine Erzeugende der zyklischen Gruppe $\Lambda\{a_P\}$ ist; jedes Element $T \in \Lambda - E$ besitzt genau eine Darstellung $T = P^n$, in welcher $n \geqq 1$ und P ein Primelement ist. Die dadurch eindeutig bestimmte natürliche Zahl $\nu(T) = n$ heiße die Vielfachheit von T.

Sind S, $T \neq E$ konjugierte Elemente von Λ, so ist offensichtlich $\nu(S) = \nu(T)$. Daher können wir für jede Klasse $\Phi \neq \{E\}$ konjugierter Elemente von Λ eindeutig definieren:

$$(3) \qquad \nu(\Phi) = \nu(T), \quad T \in \Phi.$$

---

[15]) Vgl. z. B. [4], pag. 29, Lemma 7 (a).

5.4. Ist Φ irgendeine Klasse konjugierter Elemente von Λ und $n \geqq 1$, so ist offenbar

(4) $$\Phi^n = \{T^n \mid T \in \Phi\}$$

wieder eine volle Klasse konjugierter Elemente. Aus 5.3 folgt nun: Jede Klasse $\Phi \neq \{E\}$ besitzt genau eine Darstellung $\Phi = \Pi^n$, in welcher $n \geqq 1$ und Π eine Primklasse ist; dabei ist $n = \nu(\Phi)$.

5.5. Wir definieren jetzt für jede Homotopieklasse $\mathfrak{W} \neq \mathfrak{O}$

(5) $$\nu(\mathfrak{W}) = \nu(\Phi_{\mathfrak{W}})\,.$$

Da die in 1.2 erklärte Zuordnung $\mathfrak{W} \to \Phi_{\mathfrak{W}}$ die Menge aller Homotopieklassen $\mathfrak{W} \neq \mathfrak{O}$ umkehrbar eindeutig auf die Menge aller Klassen $\Phi \neq \{E\}$ konjugierter Elemente von Λ abbildet und offensichtlich die Eigenschaft

(6) $$\Phi_{\mathfrak{W}^n} = \Phi^n_{\mathfrak{W}}\,, \quad n \geqq 1$$

besitzt, so folgt aus 5.4: Eine Homotopieklasse $\mathfrak{P}$ ist genau dann primitiv im Sinne von 1.1, wenn $\nu(\mathfrak{P}) = 1$, i. e. wenn $\Phi_{\mathfrak{P}}$ eine Primklasse ist. Jede Homotopieklasse $\mathfrak{W} \neq \mathfrak{O}$ besitzt genau eine Darstellung $\mathfrak{W} = \mathfrak{P}^n$, in welcher $n \geqq 1$ und $\mathfrak{P}$ primitiv ist; dabei ist $n = \nu(\mathfrak{W})$.

5.6. Es sei $A = \{a_T \mid T \in \Lambda - E\}$ die Menge aller Achsen von Λ. Wegen (2) wirkt Λ als Permutationsgruppe auf A. Wir heißen nun zwei Achsen $a_1, a_2 \in A$ äquivalent, wenn es ein solches $S \in \Lambda$ gibt, daß $a_2 = S(a_1)$. Da dieser Äquivalenzbegriff offensichtlich symmetrisch, reflexiv und transitiv ist, bewirkt er eine Einteilung von A in Äquivalenzklassen.

5.7. Es sei $a_n \in A$, $n = 1, 2, \ldots$; ein Repräsentantensystem dieser Klasseneinteilung von A, und es sei P_n eine Erzeugende der unendlichzyklischen Gruppe $\Lambda\{a_n\}$. Dann ist $\{P_n^m \mid n = 1, 2, \ldots ;\, m \neq 0\}$ ein Repräsentantensystem aller Klassen $\Phi \neq \{E\}$ konjugierter Elemente von Λ.

5.8. Beweis: Wir haben zu zeigen:

(a) Zu jedem $T \in \Lambda - E$ gibt es einen Index n, eine ganze Zahl $m \neq 0$ und ein $S \in \Lambda$ derart, daß $STS^{-1} = P_n^m$.

(b) Aus $P_{n_1}^{m_1} = SP_{n_2}^{m_2}S^{-1}$, $m_1, m_2 \neq 0$, $S \in \Lambda$ folgt: $n_1 = n_2$, $m_1 = m_2$.

ad (a): Offenbar gibt es einen Index n und ein $S \in \Lambda$ so, daß $S(a_T) = a_n$. Dann ist nach (2) $a_{STS^{-1}} = a_n$, i. e. $STS^{-1} \in \Lambda\{a_n\} = [P_n]$. Somit gibt es in der Tat ein solches $m \neq 0$, daß $STS^{-1} = P_n^m$.

ad (b): Aus

(7) $$P_{n_1}^{m_1} = SP_{n_2}^{m_2}S^{-1}$$

folgt nach (1) und (2): $a_{P_{n_1}} = S(a_{P_{n_2}})$, i. e. $a_{n_1} = S(a_{n_2})$. Dann ist aber $n_1 = n_2$ und somit $a_{n_1} = S(a_{n_1})$, i. e. $S \in \Lambda\{a_{n_1}\} = [P_{n_1}]$. Somit folgt wegen $n_1 = n_2$ aus (7): $P_{n_1}^{m_1} = P_{n_1}^{m_2}$. Folglich ist $m_1 = m_2$, da ja P_{n_1} ein Element unendlicher Ordnung ist. q. e. d.

5.9. Es sei

(8) $$\Lambda = \bigcup_j \Lambda\{a_n\}\,T_{nj}$$

die Restklassenzerlegung von Λ nach der zyklischen Untergruppe $\Lambda\{a_n\} = [P_n]$. Dann gilt:

(a) Zu jedem Element $T \in \Lambda - E$ gibt es zwei Indices n, j und eine ganze Zahl $m \neq 0$ derart, daß $T = T_{nj}^{-1} P_n^m T_{nj}$.

(b) Aus $T_{n_1 j_1}^{-1} P_{n_1}^{m_1} T_{n_1 j_1} = T_{n_2 j_2}^{-1} P_{n_2}^{m_2} T_{n_2 j_2}$, $m_1, m_2 \neq 0$, folgt: $n_1 = n_2$, $j_1 = j_2$, $m_1 = m_2$.

5.10. Beweis von (a): Nach 5.7 gibt es zu jedem $T \in \Lambda - E$ einen Index n, eine ganze Zahl $m \neq 0$ und ein $S \in \Lambda$ derart, daß

$$T = S^{-1} P_n^m S \,. \tag{9}$$

Zu diesem S und n gibt es nach (8) einen Index j und einen Exponenten l derart, daß $S = P_n^l T_{nj}$. Daraus und aus (9) folgt aber $T = T_{nj}^{-1} P_n^m T_{nj}$. q. e. d.

Beweis von (b): Aus der Voraussetzung folgt

$$P_{n_1}^{m_1} = S P_{n_2}^{m_2} S^{-1}\,, \quad S = T_{n_1 j_1} T_{n_2 j_2}^{-1} \in \Lambda\,, \quad m_1, m_2 \neq 0\,.$$

Daher ist nach 5.7 $n_1 = n_2$, $m_1 = m_2$ und somit

$$P_{n_1}^{m_1} = S P_{n_1}^{m_1} S^{-1}\,, \quad m_1 \neq 0 \tag{10}$$

$$S = T_{n_1 j_1} T_{n_1 j_2}^{-1}\,. \tag{11}$$

Aus (10) folgt nach (1) und (2): $a_{P_{n_1}} = S(a_{P_{n_1}})$, i. e. $a_{n_1} = S(a_{n_1})$. Somit ist $S \in \Lambda\{a_{n_1}\}$. Daraus und aus (11) folgt aber wegen (8): $j_1 = j_2$. q. e. d.

5.11. Weil das Fundamentalpolygon $\mathfrak{B}$ von Λ kompakt ist, folgt aus 3.7 und 5.9 für $\sigma > 1$:

$$\begin{aligned} \iint G(p,p,s)\,d\omega_p - 4\pi(g-1) &= \sum_{m\neq 0} \sum_n \sum_j \iint_{\mathfrak{B}} \operatorname{Cos}^{-s} \varrho(T_{nj}^{-1} P_n^m T_{nj} p, p)\, d\omega_p \\ &= \sum_{m\neq 0} \sum_n \sum_j \iint_{\mathfrak{B}} \operatorname{Cos}^{-s} \varrho(P_n^m T_{nj} p, T_{nj} p)\, d\omega_p \\ &= \sum_{m\neq 0} \sum_n \sum_j \iint_{T_{nj}(\mathfrak{B})} \operatorname{Cos}^{-s} \varrho(P_n^m p, p)\, d\omega_p\,. \end{aligned} \tag{12}$$

Hieraus folgt für alle Indices n und alle $m \neq 0$

$$\sum_j \iint_{T_{nj}(\mathfrak{B})} \operatorname{Cos}^{-\sigma} \varrho(P_n^m p, p)\, d\omega_p < \iint_{\mathfrak{B}} G(p,p,\sigma)\, d\omega_p < \infty \quad \text{für} \quad \sigma > 1\,. \tag{13}$$

Wir definieren nun

$$\mathfrak{B}_n = \bigcup_j T_{nj}(\mathfrak{B}). \tag{14}$$

$\mathfrak{B}_n$ ist Vereinigungsmenge der paarweise nicht überlappenden Fundamentalpolygone $T_{nj}(\mathfrak{B})$. Daher folgt aus (13), daß die Funktion $\operatorname{Cos}^{-s} \varrho(P_n^m p, p)$ für $\sigma > 1$ absolut integrierbar über $\mathfrak{B}_n$ ist und daß

$$\sum_j \iint_{T_{nj}(\mathfrak{B})} \operatorname{Cos}^{-s} \varrho(P_n^m p, p)\, d\omega_p = \iint_{\mathfrak{B}_n} \operatorname{Cos}^{-s} \varrho(P_n^m p, p)\, d\omega_p\,, \quad \sigma > 1\,, \quad m \neq 0\,.$$

Somit folgt aus (12)

$$\iint_{\mathfrak{B}} G(p,p,s)\, d\omega - 4\pi(g_{\mathfrak{F}} - 1) = \sum_{m\neq 0} \sum_n I_{nm}(s)\,, \quad \sigma > 1 \tag{15}$$

mit

$$I_{nm}(s) = \iint_{\mathfrak{B}_n} \operatorname{Cos}^{-s} \varrho(P_n^m p, p)\, d\omega_p\,, \quad m \neq 0\,, \quad \sigma > 1\,. \tag{16}$$

5.12. Zur Berechnung von $I_{nm}(s)$ für festes n können wir nach 2.6 auf $\mathfrak{H}$ ein Poincaréschen Koordinatensystem

$$z(p) = x_1(p) + i\,x_2(p)\,, \quad -\infty < x_1(p) < +\infty\,, \quad x_2(p) > 0$$

in solcher Weise einführen, daß

(17) $$z(P_n p) = \alpha z(p)\,, \quad \alpha > 1\,.$$

Dann gilt nach 2.6

(18) $$\operatorname{Cos}\varrho(P_n^m p, p) = 1 + (\operatorname{Cos}\mu(P_n^m) - 1)\left(1 + \frac{x_1^2(p)}{x_2^2(p)}\right)$$

(19) $$\log\alpha = \mu(P_n) = \frac{\mu(P_n^m)}{|m|}\,.$$

Da $[P_n^{-1}] = [P_n] = \Lambda\{a_n\} = \Lambda\{a_{P_n}\} = \Lambda\{a_{P_n^{-1}}\}$, so sind P_n und P_n^{-1} nach 5.3 Primelemente, und es gilt daher $\nu(P_n^m) = |m|$. Somit folgt aus (19):

(20) $$\log\alpha = \frac{\mu(P_n^m)}{\nu(P_n^m)}\,.$$

5.13. Aus (8) und (14) folgt leicht, daß $\mathfrak{B}_n$ ein Fundamentalbereich der zyklischen Gruppe $\Lambda\{a_n\} = [P_n]$ ist. Nun ist aber der Integrand von $I_{nm}(s)$ offensichtlich automorph bezüglich dieser Gruppe. Daher ändert sich der Wert des Integrals nicht, wenn wir den Integrationsbereich $\mathfrak{B}_n$ durch irgendeinen andern (meßbaren) Fundamentalbereich der Gruppe $\Lambda\{a_n\} = [P_n]$ ersetzen. Wegen (17) ist offenbar

(21) $$\mathfrak{B}_n^* = \{p\,|-\infty < x_1(p) < +\infty,\ 1 \leqq x_2(p) < \alpha\}$$

ein solcher Fundamentalbereich. Somit gilt

(22) $$I_{nm}(s) = \iint\limits_{\mathfrak{B}_n^*} \operatorname{Cos}^{-s}\varrho(P_n^m p, p)\,d\omega_p\,, \quad \sigma > 1\,, \quad m \neq 0\,.$$

Beachten wir noch, daß nach 2.1 (1) $d\omega_p = x_2^{-2}dx_1dx_2$, so folgt aus (18), (21) und (22)

(23) $$I_{nm}(s) = \int\limits_1^\alpha J_{nm}(x_2, s)\,x_2^{-2}dx_2\,, \quad \sigma > 1\,, \quad m \neq 0$$

mit

$$J_{nm}(x_2, s) = \int\limits_{-\infty}^{+\infty}\left\{1 + (\operatorname{Cos}\mu(P_n^m) - 1)\left(1 + \frac{x_1^2}{x_2^2}\right)\right\}^{-s} dx_1\,, \quad x_2 > 0\,.$$

Führen wir im Integral J_{nm} die Substitution $x_1 = x_2 x$ aus, so kommt

$$J_{nm}(x_2, s) = x_2 \operatorname{Cos}^{-s}\mu(P_n^m)\int\limits_{-\infty}^{+\infty}\left(1 + \frac{\operatorname{Cos}\mu(P_n^m) - 1}{\operatorname{Cos}\mu(P_n^m)}\,x^2\right)^{-s} dx\,.$$

Führen wir endlich noch die Substitution

$$x = y\left(\frac{\operatorname{Cos}\mu(P_n^m)}{\operatorname{Cos}\mu(P_n^m) - 1}\right)^{\frac{1}{2}}$$

aus, so erhalten wir unter Verwendung einer Formel aus der Theorie der

Gammafunktion[16])

$$J_{nm}(x_2, s) = x_2\left(\frac{\operatorname{Cos}\mu(P_n^m)}{\operatorname{Cos}\mu(P_n^m)-1}\right)^{\frac{1}{2}}\operatorname{Cos}^{-s}\mu(P_n^m)\int_{-\infty}^{+\infty}(1+y^2)^{-s}\,dy$$
$$= x_2\left(\frac{\operatorname{Cos}\mu(P_n^m)}{\operatorname{Cos}\mu(P_n^m)-1}\right)^{\frac{1}{2}}\operatorname{Cos}^{-s}\mu(P_n^m)\sqrt{\pi}\,\frac{\Gamma(s-\frac{1}{2})}{\Gamma(s)}\,.$$

Hieraus und aus (23) folgt unter Berücksichtigung von (20)

$$I_{nm}(s) = \frac{\mu(P_n^m)}{\nu(P_n^m)}\left(\frac{\operatorname{Cos}\mu(P_n^m)}{\operatorname{Cos}\mu(P_n^m)-1}\right)^{\frac{1}{2}}\operatorname{Cos}^{-s}\mu(P_n^m)\sqrt{\pi}\,\frac{\Gamma(s-\frac{1}{2})}{\Gamma(s)}\,,\quad m \neq 0\,,\quad \sigma > 1\,.$$

Somit folgt aus (15)

$$\frac{1}{\sqrt{\pi}}\,\frac{\Gamma(s)}{\Gamma(s-\frac{1}{2})}\iint_{\mathfrak{B}} G(p,p;s)\,d\omega - 4\sqrt{\pi}\,(g_{\mathfrak{F}}-1)\frac{\Gamma(s)}{\Gamma(s-\frac{1}{2})}$$
$$= \sum_{m\neq 0}\sum_{n}\frac{\mu(P_n^m)}{\nu(P_n^m)}\left(\frac{\operatorname{Cos}\mu(P_n^m)}{\operatorname{Cos}\mu(P_n^m)-1}\right)^{\frac{1}{2}}\operatorname{Cos}^{-s}\mu(P_n^m)\,.$$

Daher gilt wegen 5.7 und (3)

$$\frac{1}{\sqrt{\pi}}\,\frac{\Gamma(s)}{\Gamma(s-\frac{1}{2})}\iint_{\mathfrak{B}} G(p,p;s)\,d\omega - 4\sqrt{\pi}\,(g_{\mathfrak{F}}-1)\frac{\Gamma(s)}{\Gamma(s-\frac{1}{2})}$$
$$= \sum_{\Phi\neq\{E\}}\frac{\mu(\Phi)}{\nu(\Phi)}\left(\frac{\operatorname{Cos}\mu(\Phi)}{\operatorname{Cos}\mu(\Phi)-1}\right)^{\frac{1}{2}}\operatorname{Cos}^{-s}\mu(\Phi)\,.$$

Wegen 1.2 (2) und 5.5 haben wir damit bewiesen:

5.14. In der Halbebene $\sigma > 1$ gilt

$$\frac{1}{\sqrt{\pi}}\,\frac{\Gamma(s)}{\Gamma(s-\frac{1}{2})}\iint_{\mathfrak{B}} G(p,p;s)\,d\omega - 4\sqrt{\pi}\,(g_{\mathfrak{F}}-1)\frac{\Gamma(s)}{\Gamma(s-\frac{1}{2})}$$
$$= \sum_{\mathfrak{W}\neq\mathfrak{O}}\frac{\mu(\mathfrak{W})}{\nu(\mathfrak{W})}\left(\frac{\operatorname{Cos}\mu(\mathfrak{W})}{\operatorname{Cos}\mu(\mathfrak{W})-1}\right)^{\frac{1}{2}}\operatorname{Cos}^{-s}\mu(\mathfrak{W})\,.$$

Literatur

[1] Bohr, H.: Zur Theorie der fast periodischen Funktionen. I. Acta math. **45**, 29—127 (1925). — [2] Doetsch, G.: Handbuch der Laplace-Transformation. I. Basel: Birkhäuser 1950. — [3] Hilbert, D.: Grundzüge einer allgemeinen Theorie der linearen Integralgleichungen. Leipzig und Berlin: B. G. Teubner 1912. — [4] Huber, H.: Über eine neue Klasse automorpher Funktionen und ein Gitterpunktproblem in der hyperbolischen Ebene. Comment. Math. Helv. **30**, 20—62 (1956). — [5] Magnus, W., u. F. Oberhettinger: Formeln und Sätze... Berlin: Springer 1943. — [6] Minakshisundaram, S., and A. Pleijel: Some properties of the eigenfunctions of the Laplace-Operator on Riemannian Manifolds. Canad. J. Math. **1**, 242—256 (1949). — [7] Seifert, H., u. W. Threlfall: Lehrbuch der Topologie. Leipzig: B. G. Teubner 1934. — [8] Weyl, H.: Die Idee der Riemannschen Fläche. 3. Aufl. Stuttgart: B. G. Teubner 1955. — [9] Wiener, N.: Tauberian Theorems. Ann. of Math. **33**, 1—100 (1932).

(Eingegangen am 15. Oktober 1958)

---

[16]) [5], pag. 4.

BRAUNER, H.
Math. Annalen 138, 27—41 (1959)

Eine Verallgemeinerung des Problems der Cesàrokurven

Von

H. BRAUNER in Wien

Wird das begleitende Dreibein einer Raumkurve längs dieser bewegt, so beschreibt jede im Dreibein festgehaltene Gerade bei dieser Bewegung eine gewisse Strahlfläche, die für alle Geraden in der rektifizierenden Ebene mit Tangentenrichtung sowie für zwei gewisse Minimalgeraden *Torsen* (also Strahlflächen von konstantem Drall null) sind. E. CESÀRO [1] hat die Frage nach solchen Kurven aufgeworfen, bei denen außer diesen genannten Geraden noch andere existieren, die bei Bewegung des begleitenden Dreibeins Torsen erzeugen. Dieses Problem führte ihn auf die sog. *Cesàrokurven*, die dadurch charakterisiert sind, daß zwischen ihrer Krümmung und Torsion eine quadratische Gleichung mit konstanten Koeffizienten und ohne absolutes Glied existiert. E. SALKOWSKI [2] hat gezeigt, daß die Zugehörigkeit zur Klasse der Cesàrokurven *nur eine notwendige Bedingung* darstellt und daß in einigen Ausnahmefällen bei Cesàrokurven keine zusätzlichen Geraden existieren, die auf Torsen führen.

Diese Fragen wurden von E. KRUPPA [3] insofern verallgemeinert, als er die Bewegung des begleitenden Dreibeins einer *windschiefen Strahlfläche* zugrunde legt. Eine andere Verallgemeinerung, die hier betrachtet werden soll, entsteht, wenn man nach allen im Dreibein einer Raumkurve festen Geraden fragt, die bei Bewegung des Dreibeins *Strahlflächen von konstantem Drall* überstreichen. Diese Fragestellung wurde durch die Entdeckung einer Reihe von Eigenschaften solcher Flächen nahegelegt [4].

Zunächst werden jene Geraden erfaßt, die bei *allen Raumkurven* auf Strahlflächen konstanten Dralls führen und dann jene Kurvenklassen genauer untersucht, wo *zusätzliche Geraden* dieser Art existieren. Diese Kurven sind *notwendig wieder Cesàrokurven*; in den einzelnen Fällen werden vor allem die interessanten Konfigurationen jener im Dreibein fester Geraden betrachtet. Diese Untersuchungen enthalten die ursprüngliche Cesàrosche Fragestellung als *Sonderfall*, für die ebenfalls eine Reihe neuer Tatsachen angeführt wird; insbesondere zeigt sich bei Aufsuchung der *hinreichenden Bedingungen*, daß SALKOWSKI eine Type von Cesàrokurven in seiner Aufzählung übersehen hat, bei denen keine zusätzlichen Geraden existieren, die bei Bewegung des Dreibeins auf Torsen führen.

1. Ist eine Raumkurve c durch den Ortsvektor $\mathfrak{x}(s)$ als Funktion ihrer Bogenlänge s gegeben und bedeuten $\mathfrak{t}, \mathfrak{h}, \mathfrak{b}$ den Tangentenvektor, Hauptnormalvektor bzw. Binormalvektor von c (die in dieser Reihenfolge ein Rechts-

system bilden mögen), so beschreibt jede *im begleitenden Dreibein* $\{\mathfrak{t}, \mathfrak{h}, \mathfrak{b}\}$ *festgehaltene Gerade* g bei Bewegung des Dreibeins längs c eine *Strahlfläche* Θ

$$\mathfrak{X}(s, v) = \mathfrak{x}(s) + p\,\mathfrak{t}(s) + q\,\mathfrak{h}(s) + r\,\mathfrak{b}(s) + v(\alpha\,\mathfrak{t}(s) + \beta\,\mathfrak{h}(s) + \gamma\,\mathfrak{b}(s)) = \tag{1}$$
$$= \mathfrak{y}(s) + v\,\mathfrak{e}(s)\,.$$

Die (konstanten) Zahlen p, q, r bestimmen dabei einen im Dreibein festen Punkt Y auf g und die (konstanten) Größen $\alpha:\beta:\gamma$ den im Dreibein festen Richtungsvektor $\mathfrak{e}$ von g. Ist $\mathfrak{e}$ ein Einsvektor bzw. ein isotroper Vektor, so gilt

$$\mathfrak{e}^2 = \alpha^2 + \beta^2 + \gamma^2 = 1 \quad \text{bzw.} \quad \mathfrak{e}^2 = \alpha^2 + \beta^2 + \gamma^2 = 0\,. \tag{2}$$

Der *Drall* d der Fläche Θ ist gegeben durch $d = (\mathfrak{e}\,\dot{\mathfrak{e}}\,\dot{\mathfrak{y}}) : \dot{\mathfrak{e}}^2$ und dafür ergibt sich nach (1) (wenn $\varkappa$ und τ die Krümmung bzw. Torsion von c ist)

$$A\varkappa^2 + B\tau^2 + C\varkappa\tau = P\varkappa + Q\tau\,. \tag{3}$$

Für die nur dem Verhältnis nach bestimmten Koeffizienten in (3) erhält man

$$A:B:C:P:Q = [\gamma(q\alpha - p\beta) - d(\alpha^2 + \beta^2)] : [\alpha(r\beta - q\gamma) - d(\beta^2 + \gamma^2)] : \tag{4}$$
$$: [\alpha(q\alpha - p\beta) + \gamma(r\beta - q\gamma) + 2\alpha\gamma d] : \alpha\gamma : -(\beta^2 + \gamma^2)\,.$$

Wir legen das Dreibein $\{\mathfrak{t}, \mathfrak{h}, \mathfrak{b}\}$ eines festen Kurvenpunktes X in die Achsen eines kartesischen Parallelkoordinatensystems (x, y, z). Die Gerade g besitzt dann die Plückerschen Strahlkoordinaten

$$p_1 : p_2 : p_3 : p_4 : p_5 : p_6 = \alpha : \beta : \gamma : (q\gamma - r\beta) : (r\alpha - p\gamma) : (p\beta - q\alpha) \tag{5}$$

und aus (4) wird

$$\varrho A = -p_3p_6 - d(p_1^2 + p_2^2)\,, \quad \varrho C = -p_1p_6 - p_3p_4 + 2d\,p_1p_3\,, \tag{6}$$
$$\varrho B = -p_1p_4 - d(p_2^2 + p_3^2)\,, \quad \varrho P = p_1p_3\,, \quad \varrho Q = -(p_2^2 + p_3^2)\,.$$

Wir fragen nun nach solchen im Dreibein von c festen Geraden, die bei Bewegung des Dreibeins längs c *Strahlflächen* Θ *von konstantem Drall* d überstreichen; solche Geraden sollen im folgenden kurz „*d-Geraden*" genannt werden. Für ihre Plückerkoordinaten erhält man aus (3) und (6)

$$\varkappa^2p_3p_6 + \tau^2p_1p_4 + \varkappa\tau(p_1p_6 + p_3p_4) + \varkappa p_1p_3 - \tau(p_2^2 + p_3^2) + \tag{7}$$
$$+ d[\varkappa^2(p_1^2 + p_2^2) + \tau^2(p_2^2 + p_3^2) - 2\varkappa\tau p_1p_3] = 0\,.$$

Bei festem d ist das ein *quadratischer Strahlkomplex* $K(d)$. Bewegt man das Koordinatensystem durch

$$x = \frac{\varkappa x' + \tau y'}{\sqrt{\varkappa^2 + \tau^2}}\,, \quad y = y' + \frac{\varkappa}{\varkappa^2 + \tau^2}\,, \quad z = \frac{-\tau x' + \varkappa z'}{\sqrt{\varkappa^2 + \tau^2}} \tag{8}$$

in eine solche Lage (x', y', z'), daß die z'-Achse in die rektifizierende Kante k fällt, der Ursprung in den Schnittpunkt der Hauptnormalen mit k und die $x'z'$-Ebene parallel zur rektifizierenden Ebene wird, so lautet der Komplex $K(d)$

$$\left(\frac{\tau}{\varkappa^2 + \tau^2} - d\right)(p_1'^2 + p_2'^2) - p_3'p_6' = 0\,. \tag{9}$$

Das ist der *Tangentenkomplex einer euklidischen Schraubung* um die rektifizierende Kante k vom Parameter $\frac{\tau}{\varkappa^2+\tau^2} - d$. Variiert d, so erhält man ∞^1 *koaxiale Schraubungen*, unter denen für $d = 0$ die oskulierende Schraubung von c vorkommt[1]).

Satz 1: *Die d-Geraden einer Kurve gehören dem Tangentenkomplex einer Schraubung um die rektifizierende Kante vom Parameter* $\frac{\tau}{\varkappa^2+\tau^2} - d$ *an.*

Um für eine gegebene Kurve c die d-Geraden aus dem Komplex $K(d)$ auszusondern, dienen die Gleichungen (3) und (6). Besteht zwischen Krümmung und Torsion von c *keine quadratische Beziehung* der Form (3), so müssen sämtliche Koeffizienten verschwinden und das liefert unter Beachtung von (2) und (6) zu jedem Wert von d zwei isotrope Lösungen

$$p_1 : \ldots : p_6 = 0 : 1 : \pm i : 0 : d : \pm i d\,, \tag{10}$$

sowie außerdem für $d = 0$

$$p_1 : \ldots : p_6 = 1 : 0 : 0 : 0 : r : 0\,. \tag{11}$$

(10) stellt bei variablem d die Minimalstrahlbüschel in den beiden Minimalebenen durch die Tangente dar, (11) alle zur Tangente parallelen Geraden der rektifizierenden Ebene. Kurven, bei denen eine quadratische Beziehung der Gestalt (3) existiert, sind als „*Cesàrokurven*" bekannt. Nur bei ihnen können außer den oben genannten d-Geraden (die sich stets ergeben) noch *zusätzliche* existieren.

Satz 2[2]): *Jede Raumkurve besitzt für jeden Wert von d zwei die Tangente normal schneidende isotrope d-Geraden, die von der Normalebene den Abstand* $\pm i d$ *und gegen die Schmiegebene die Böschung* $\pm i$ *besitzen; außerdem sind alle zur Tangente parallelen Geraden der rektifizierenden Ebene* 0*-Geraden. Die einzigen Kurven, bei denen noch zusätzliche d-Geraden existieren können, sind die Cesàrokurven.*

2. Im folgenden sollen die bei Cesàrokurven *zusätzlich* existierenden d-Geraden untersucht werden und als „*eigentliche d-Geraden*" bezeichnet werden. Ob (3) für die Existenz eigentlicher d-Geraden auch hinreichend ist, muß erst geklärt werden.

Sei nun c *eine Cesàrokurve.* Die sie bestimmende Gleichung (3) ist dann reduzibel, wenn die rechte Seite der Gleichung ein Teiler der linken ist. Das ist der Fall, wenn

$$\Delta \equiv A Q^2 - C P Q + B P^2 = 0 \tag{12}$$

gilt und (3) nimmt dann die Gestalt an

$$(P\varkappa + Q\tau)(a\varkappa + b\tau - 1) = 0 \tag{13}$$

mit

$$A = Pa\,, \quad B = Qb\,, \quad C = Pb + Qa\,. \tag{14}$$

---

[1]) Diese Tatsache verwendet E. Kruppa [3] im Falle $d = 0$, um die Geraden zu finden, die bei Bewegung des Dreibeins auf Torsen führen.

[2]) Diesen Satz habe ich bereits in [4] abgeleitet.

c besteht dann aus einer Böschungslinie und einer Bertrandkurve. Wir setzen c zunächst als *nicht reduzibel* ($\Delta \neq 0$) voraus.

Wir untersuchen die *eigentlichen d-Geraden einer nicht reduziblen Cesàrokurve*, für die $P \neq 0$ gilt („*allgemeiner Fall*"). Infolge (6) ist mit $P \neq 0$ auch $p_1 \neq 0$, $p_3 \neq 0$ und man erhält aus (6)

$$\begin{aligned} &K_1 \ldots A p_1 p_3 + P p_3 p_6 + d P(p_1^2 + p_2^2) = 0\,, \\ (15)\qquad &K_2 \ldots B p_1 p_3 + P p_1 p_4 + d P(p_2^2 + p_3^2) = 0\,, \\ &K_3 \ldots P(p_2^2 + p_3^2) + Q p_1 p_3 = 0\,. \end{aligned}$$

Jede dieser Gleichungen stellt einen *quadratischen Strahlkomplex* dar. Eine Untersuchung zeigt, daß (für $d \neq 0$) K_1 der Tangentenkomplex einer euklidischen Schraubung vom Parameter d um die zur Binormale parallele Gerade $x = 0$, $y = -A : P$ der Normalebene ist, K_2 der Tangentenkomplex einer euklidischen Schraubung vom Parameter d um die zur Tangente parallele Gerade $z = 0$, $y = B : P$ der Schmiegebene und K_3 der Treffgeradenkomplex eines Fernkegelschnittes. Die Komplexkegel von K_3 sind orthogonale Kegel; sie werden von den Parallelebenen zur Normalebene in Kreisen geschnitten, die die Schmiegebene in Punkten der Tangente von c berühren. Die eigentlichen d-Geraden von c sind die gemeinsamen Strahlen dieser drei quadratischen Komplexe mit dem Komplex $K(d)$. Aus K_1 erhält man p_6 zu

$$p_6 = -\frac{A}{P}\,p_1 - \frac{d}{p_3}\,(p_1^2 + p_2^2)\,, \tag{16}$$

weiters folgt aus K_2 und K_3

$$P p_4 = (d Q - B) p_3\,. \tag{17}$$

Die eigentlichen d-Geraden gehören also einem *Gebüsch um die zur Tangente im Abstand* $(d Q - B) : P$ *parallele Gerade der Schmiegebene* an. Aus (6) erhält man ferner die Beziehung

$$P(p_1 p_6 + p_3 p_4) + (C - 2 d P) p_1 p_3 = 0\,. \tag{18}$$

Setzt man hier (16) und (17) ein und eliminiert p_2 unter Verwendung von K_3, so erhält man für

$$\lambda = p_1 : p_3 \tag{19}$$

die Gleichung

$$d P \lambda^3 + (A - d Q)\,\lambda^2 + (P d - C)\,\lambda + B - d Q = 0\,. \tag{20}$$

Denken wir uns im folgenden die Lösungen dieser kubischen Gleichung bestimmt, so erhält man unter Verwendung von (6), (17), (18), (19) für die *Plückerkoordinaten der eigentlichen d-Geraden*

$$\begin{aligned} p_1 : p_2 : p_3 : p_4 : p_5 : p_6 = \lambda : \sqrt{-\frac{Q\lambda + P}{P}} : 1 : \frac{dQ - B}{P} : \\ (21)\qquad : \frac{1}{P}\sqrt{-\frac{P}{Q\lambda + P}}\,[d P \lambda^2 + (A + B - 2 d Q)\,\lambda - d P] : \\ : \frac{-1}{P}\,[d P \lambda^2 + (A - d Q)\,\lambda - d P]\,. \end{aligned}$$

Das sind für $d \neq 0$ *sechs* Geraden, die paarweise zur Hauptnormalen symmetrisch liegen und die Schmiegebene in Punkten der zur Tangente parallelen Geraden $z = 0$, $y = (dQ - B) : P$ durchsetzen. Ist speziell $d = 0$, so reduziert sich (20) zu

$$A\lambda^2 - C\lambda + B = 0 \tag{22}$$

und dies führt unter Verwendung von (21) auf *vier* 0-Geraden[3]).

(21) liefert nur dann keine Geraden, wenn sämtliche Plückerkoordinaten verschwinden. Das ist für $P \neq 0$ nur dann möglich, wenn $Q\lambda + P = 0$ und damit wegen des Komplexes K_3 auch $p_2 = 0$ gilt. In diesem Fall erhält man nämlich aus (6)

$$d(P^2 + Q^2) - Q(A + B) = 0 \tag{23}$$

und daher ist auch $dP\lambda^2 + (A + B - 2dQ)\lambda - dP = 0$. Andererseits folgt aus (20) mit $\lambda = -P : Q$

$$d(P^2 + Q^2)^2 - Q(AP^2 + CPQ + BQ^2) = 0\,. \tag{24}$$

(23) und (24) sind nur dann vereinbar, wenn $\Delta = AQ^2 - CPQ + BP^2 = 0$ gilt, die Cesàrokurve also *reduzibel* ist.

Satz 3: *Zur rektifizierenden Ebene parallele eigentliche d-Geraden können nur bei zerfallenden Cesàrokurven existieren.*

Wir fragen nun speziell nach *isotropen eigentlichen d-Geraden*. Unter Berücksichtigung von (2) folgt aus (21)

$$\lambda\left(\lambda - \frac{Q}{P}\right) = 0 \tag{25}$$

und wegen $P \neq 0$ ist $\lambda \neq 0$ und daher $P\lambda - Q = 0$. Dies liefert eingesetzt in (20)

$$P(AQ^2 - CPQ + BP^2) = P\Delta = 0\,, \tag{26}$$

d. h. die Cesàrokurve zerfällt.

Satz 4: *Bei einer nicht reduziblen Cesàrokurve mit $P \neq 0$ existieren keine isotropen eigentlichen d-Geraden. Zu jedem Wert von $d \neq 0$ gibt es sechs (reelle oder komplexe) eigentliche d-Geraden, für $d = 0$ vier.*

Läßt man *d variieren*, so erhält man eine von allen eigentlichen *d*-Geraden gebildete, mit dem Dreibein fest verbundene *Strahlfläche* Ω. Die *Spurkurve* dieser Strahlfläche Ω in der Schmiegebene ($z = 0$) ergibt sich aus (21) zu

$$x = \frac{-1}{P}\sqrt{-\frac{P}{Q\lambda + P}}\,[dP\lambda^2 + (A + B - 2dQ)\lambda - dP]\,, \quad y = \frac{dQ - B}{P}\,, \tag{27}$$

wobei für die Parameter λ und d nach (20) gilt

$$d = \frac{-A\lambda^2 + C\lambda - B}{(\lambda^2 + 1)(P\lambda - Q)}\,. \tag{28}$$

Aus (27) und (28) erhält man unter Verwendung des rationalisierenden Parameters

$$t = \sqrt{-\frac{P}{Q\lambda + P}} \tag{29}$$

---

[3]) Die für $d = 0$ gültige Gleichung (22) findet sich schon bei SALKOWSKI [2]. Eine explizite Darstellung der Lösungsgeraden für $d = 0$ [analog zur Darstellung (21)] dürfte bisher in der Literatur nicht vorkommen.

für die Spurkurve von Ω in homogenen Koordinaten $x = x_1 : x_0$, $y = x_2 : x_0$, $z = x_3 : x_0$ eine Gleichung der Bauart

$$x_0 : x_1 : x_2 : x_3 = t\mathfrak{P}^3(t^2) : \mathfrak{P}^4(t^2) : t(t^2+1)\,\mathfrak{P}^2(t^2) : 0\,, \tag{30}$$

wobei $\mathfrak{P}^k(t^2)$ ein Polynom k-ten Grades in t^2 bedeutet. Diese zur Hauptnormale symmetrische, rationale *Kurve* 8. *Ordnung* besitzt für $t = 0$ und $t = \infty$ im Fernpunkt der Tangente von c je einen einfachen Punkt; die restlichen sechs Fernpunkte errechnen sich aus $\mathfrak{P}^3(t^2) = 0$.

Die *Strahlfläche* Ω ist durch eine Korrespondenz zwischen dieser Kurve 8. Ordnung und dem zum Komplex K_3 gehörigen Fernkegelschnitt

$$x_0 : x_1 : x_2 : x_3 = 0 : -P(1+t^2) : Qt : Qt^2 \tag{31}$$

bestimmt, die, wie sich aus (30), (29), (20) ergibt, eine (1,1)-Korrespondenz ist, in der der Fernpunkt der Tangente von c belegt mit dem Parameterwert $t = 0$ selbstentsprechend ist. Das Erzeugnis dieser Korrespondenz ist daher eine *Strahlfläche 9. Grades*, die zur Hauptnormale axial symmetrisch ist und die Schmiegebene außer in der obigen Kurve *8.* Ordnung noch in dem Tangente von c schneidet. Ihre Fernkurve ist der Fernkegelschnitt (31), die Ferngerade der rektifizierenden Ebene und weitere sechs Ferngeraden.

Satz 5: *Die eigentlichen d-Geraden einer nicht reduziblen Cesàrokurve mit* $P \neq 0$ *erfüllen eine zur Hauptnormale axial symmetrische, rationale Strahlfläche 9. Grades.*

3. Nun ist noch die Diskussion jener *nicht zerfallenden Cesàrokurven* ausständig, *für die* $P = 0$ *gilt*. Wegen (6) ist dann $p_1 p_3 = 0$, d. h. die eigentlichen d-Geraden sind entweder zur Schmiegebene oder zur Normalebene oder zur Hauptnormale parallel.

Es sei zunächst $P = p_1 = 0$, $p_3 \neq 0$. Aus (6) folgt dann

$$d = B : Q\,, \tag{32}$$

d. h. alle eigentlichen d-Geraden führen auf Strahlflächen vom *gleichen konstanten Drall* $B : Q$. Aus (6) folgt ferner die Parameterdarstellung

$$\begin{aligned} p_1 : \ldots : p_6 = 0 : Q\mu^2 : Q\mu : C\mu(\mu^2+1) : -[\mu^2(A-B)+A] :\\ :\mu\,[\mu^2(A-B)+A]\,. \end{aligned} \tag{33}$$

Die eigentlichen d-Geraden erfüllen daher eine *konoidale Strahlfläche 3. Grades* Φ, deren Gleichung nach (33) lautet

$$[AC-(A-B)(C-Qy)]^2 + Q^2[Cx+(A-B)z]\,[x(C-Qy)+Az] = 0\,. \tag{34}$$

Φ besitzt die Ferngerade der Normalebene als einfache Leitgerade und die die Hauptnormale im Punkt $y = BC : Q(B-A)$ normal schneidende Gerade der Böschung $C : (B-A)$ gegen die Schmiegebene zur doppelten Leitgeraden. Φ schneidet die rektifizierende Ebene in ihrer Ferngeraden und einer Hyperbel, die den Ursprung als Mitte und eine zur doppelten Leitgeraden parallele Asymptote besitzt. Φ besteht daher aus allen im hyperbolischen Netz der beiden Leitgeraden enthaltenen Strahlen, die diese Hyperbel treffen; sie ist von 8. Art ([5] S. 196) und für $A(B-A) < 0$ vom *I*. Typ, für $A(B-A) > 0$

vom *II.* Typ. Ihr Normalriß auf die Normalebene ist (übrigens unabhängig vom Wert von B) eine Parabel, deren Brennpunkt im Ursprung und deren Scheitel auf der Hauptnormale im Punkt $y = C : Q$ liegt. In (33) kommen für $Q^2\mu^2(\mu^2+1) = 0$ isotrope d-Geraden vor; wegen $c \neq 0$ ist $\mu \neq 0$, $\mu^2 + 1 = 0$ aber führt auf die isotropen Geraden (5), d. h. es gibt *keine eigentlichen isotropen d-Geraden.*

Wir betrachten einige *Sonderfälle,* in denen c nicht reduzibel wird. Ist $C = 0$, so wird Φ ein *gerades Konoid* mit der Tangente als doppelte Leitgerade. Ist $A = B(\neq 0)$, so ist Φ eine *Cayleysche Strahlfläche* 3. Grades von 11. Art ([5], S. 196), die in einem parabolischen Netz um die Ferngerade der Normalebene enthalten ist. Die Ferngerade der rektifizierenden Ebene ist Fernerzeugende, die Striktionslinie speziell eine kubische hyperbolische Parabel (vgl. [5], S. 200). Ist außer $A = B$ noch $C = 0$, so erfüllen die eigentlichen d-Geraden jene *Strahlschar 2. Grades* auf dem gleichseitigen hyperbolischen Paraboloid

$$Az - Qxy = 0\,, \tag{35}$$

die die Hauptnormale enthält. Ist insbesondere $B = 0$, so sind nach (34) alle eigentlichen d-Geraden speziell 0-Geraden und c ist eine Kurve der von J. ANDRADE untersuchten Art [6], ([2], S. 537)

$$A\varkappa^2 + C\varkappa\tau = Q\tau\,. \tag{36}$$

Φ lautet in diesem Fall

$$A^2y^2 + (Cx + Az)^2 - Qxy(Cx + Az) = 0\,; \tag{37}$$

sie schneidet die rektifizierende Ebene in ihrer doppelten Leitgeraden, ist vom *I.* Typ und kann konstruktiv unter Verwendung der Umrißparabel auf die Normalebene erzeugt werden. Ist außer $B = 0$ noch $C = 0$, so ist die *Torsion von c proportional zum Quadrat der Krümmung.* Φ besitzt dann jene metrisch spezielle Form, bei der die drei Ferngeraden ein Poldreieck des absoluten Kegelschnittes bilden (vgl. [5], S. 199).

Nun untersuchen wir den Fall $P = p_3 = 0$, $p_1 \neq 0$. Aus (6) folgt

$$d = A : Q(1 + \mu^2) \tag{38}$$

und die Parameterdarstellung

$$\begin{aligned} p_1 : \ldots : p_6 = {} & Q\mu^2(1+\mu^2) : Q\mu(1+\mu^2) : 0 : [B(1+\mu^2) - A] : \\ & : -\mu[B(1+\mu^2) - A] : C(1+\mu^2)\,. \end{aligned} \tag{39}$$

Das ist eine *konoidale Strahlfläche 4. Grades* Ψ mit der Darstellung

$$x = \frac{C}{Q\mu} + v\frac{\mu}{\sqrt{1+\mu^2}}\,, \quad y = v\frac{1}{\sqrt{1+\mu^2}}\,, \quad z = \frac{A - B(1+\mu^2)}{Q\mu(1+\mu^2)} \tag{40}$$

und der Gleichung

$$\begin{aligned} & Q^3z(x^2+y^2)\,[(B-A)x + Cz] + Q^2y[Cxz(3A-2B) - 2C^2z^2 + \\ & + (B-A)^2y^2 + (B-A)Bx^2] + CQ[C^2z^2 + \\ & + BCxz - 2B(B-A)y^2] + C^2B^2y = 0\,. \end{aligned} \tag{41}$$

Zu jedem Wert von d gehören zwei zur Hauptnormale symmetrische Erzeugenden parallel zur Schmiegebene. Der Umriß dieser Fläche auf die Schmiegebene ist die Parabel

$$Qx^2 - 4Cy = 0\,. \tag{42}$$

Ψ besteht aus allen Sehnen der Raumkubik k

$$x_0 : x_1 : x_2 : x_3 = Q(t^2+1) : \frac{AC}{B-A}\,t : \left[C\,t^2 + \frac{BC}{B-A}\right] : t[(A-B)t^2 - B]\,, \tag{43}$$

die die Ferngerade der Schmiegebene treffen; sie ist daher nach der Sturmschen Einteilung von IV. Art ([5], S. 255). k ist eine zur Hauptnormale symmetrische kubische Ellipse durch den Fernpunkt der Binormale, deren Scheitel der Punkt $(0, -CB : Q(A-B), 0)$ und deren Normalriß auf die Schmiegebene der Kreis

$$Q^2(B-A)(x^2+y^2) - QC(2B-A)y + BC^2 = 0 \tag{44}$$

ist. Aus (39) folgt, daß es *keine isotropen eigentlichen d-Geraden* gibt.

Wir betrachten einige *Sonderfälle*, in denen c nicht reduzibel wird. Ist $A = B(\neq 0)$, so wird die Doppelkurve von Ψ die gleichseitige Hyperbel

$$Q^2xz + AC = 0\,, \quad y = C : Q\,; \tag{45}$$

sie schneidet die Fernleitlinie der Fläche im Fernpunkt der Tangente von c. Diese Fernleitgerade ist nun selbst Flächenerzeugende ($\mu = 0$), und zwar Torsallinie zweiter Ordnung. Die zugehörige Torsalebene ist die Schmiegebene. Die Strahlfläche ist daher von VI. Art ([5], S. 264). Ist $C = 0$, so ist Ψ ein *gerades Konoid* mit der Binormale als dreifache Leitgerade. Seine Gleichung lautet nach (41)

$$Qxz(x^2+y^2) + y[B(x^2+y^2) - Ay^2] = 0\,. \tag{46}$$

Dieses Konoid wird durch eine (1,3)-Korrespondenz zwischen der Binormale und der Ferngeraden der Schmiegebene erzeugt und ist daher vom X. Typ ([5], S. 265). Eine Fläche derselben Art entsteht für $B = C = 0$, also für eine Kurve c, deren *Torsion proportional dem Quadrat der Krümmung* ist. Ist außer $C = 0$ speziell $A = B(\neq 0)$, so wird aus (46)

$$Qz(x^2+y^2) + Axy = 0\,. \tag{47}$$

Das ist ein *Plückerkonoid* von der Höhe $-A : Q$, dessen Mittelerzeugenden in der Tangente und Hauptnormale liegen. Die *oskulierenden Schraubungen* einer solchen Kurve c besitzen den *konstanten Parameter* $A : Q$.

Ist nun $P = p_1 = p_3 = 0$, so folgt aus (6)

$$\varrho A = -dp_2^2\,, \quad \varrho B = -dp_2^2\,, \quad C = 0\,, \quad \varrho Q = -p_2^2\,, \tag{48}$$

d. h. es liegt eine Cesàrokurve der Bauart $A(\varkappa^2+\tau^2) = Q\tau$ vor, deren *oskulierende Schraubungen* sämtlich den *konstanten Parameter* $A : Q$ besitzen. In diesem Fall sind alle ∞^2 zur Hauptnormale parallelen Geraden eigentliche d-Geraden zum Wert $d = A : Q$.

Satz 6: *Bei nicht reduziblen Cesàrokurven mit $P = 0$ gibt es keine eigentlichen isotropen d-Geraden. Zu jedem Wert von $d \neq 0$ existieren zwei zur Hauptnormale*

symmetrische und zur Schmiegebene parallele eigentliche d-Geraden, die in ihrer Gesamtheit eine konoidale Strahlfläche 4. Grades bilden. Für $d = B : Q$ existieren außerdem ∞^1 zur Normalebene parallele eigentliche d-Geraden, die eine konoidale Strahlfläche 3. Grades erfüllen. Speziell bei den Kurven $A = B$, $C = 0$, deren oskulierende Schraubungen den konstanten Parameter $A : Q$ besitzen, erfüllen alle eigentlichen d-Geraden für $d \neq A : Q$ ein Plückerkonoid um die Binormale mit der Schmiegebene als Mittelebene; für $d = A : Q$ sind alle ∞^2 zur Hauptnormale parallelen Geraden sowie eine zur Normalebene parallele Strahlschar 2. Grades auf einem gleichseitigen hyperbolischen Paraboloid eigentliche d-Geraden.

4. c sei nun eine *reduzible Cesàrokurve*; nach (13) gilt dann

$$(P\varkappa + Q\tau)(a\varkappa + b\tau - 1) = 0 .$$

Wir untersuchen zunächst den „*allgemeinen Fall*" $P \neq 0$. Für die eigentlichen d-Geraden gilt (20) und nach (13) daher

$$(P\lambda - Q)\,[a\lambda - b + d(\lambda^2 + 1)] = 0 , \tag{49}$$

wobei die Bezeichnung (19) verwendet wird. Analog zu (21) erhält man

$$\begin{aligned} p_1 : p_2 : p_3 : p_4 : p_5 : p_6 = \lambda : \sqrt{-\frac{Q\lambda + P}{P}} : 1 : \frac{Q}{P}(d - b) : \\ : \sqrt{-\frac{P}{Q\lambda + P}}\left[d\lambda^2 - \left(2d\frac{Q}{P} - a - b\frac{Q}{P}\right)\lambda - d\right] : \\ : \left[d(1 - \lambda^2) - \left(a - d\frac{P}{Q}\right)\lambda\right]. \end{aligned} \tag{50}$$

Diese Darstellung versagt nur, wenn $p_2 = Q\lambda + P = 0$ gilt, also zur rektifizierenden Ebene parallele eigentliche d-Geraden vorliegen. (50) liefert dann isotrope Geraden, wenn $\lambda^2 - \frac{Q\lambda + P}{P} + 1 = 0$ gilt und daher wegen $P \neq 0, \lambda \neq 0$

$$\lambda = Q : P . \tag{51}$$

Nach (13) besteht c aus einer *Böschungslinie* und einer *Bertrandkurve*, wobei die erstere speziell eine *ebene Kurve* ($P = 0$), die letztere ein *windschiefer Kreis* ($b = 0$) oder eine *Kurve konstanter Torsion* ($a = 0$) sein kann. Man erhält die eigentlichen d-Geraden dieser beiden Kurvengattungen aus (50), wenn man im Falle der Böschungslinien $P : Q$ konstant und a und b variabel, im Falle der Bertrandkurven a und b konstant und $P : Q$ variabel setzt.

Wir betrachten zunächst die *Böschungslinien* und setzen

$$a = \mu_1 , \quad b = \mu_2 . \tag{52}$$

Aus (50) folgt, daß sämtliche d-Geraden eine *Strahlkongruenz* bilden, die im Strahlgebüsch mit der Achse q

$$q_1 : \ldots : q_6 = \frac{P}{Q} : 0 : -1 : \frac{Q}{P}d : 0 : d \tag{53}$$

enthalten ist. Diese Achse q schneidet die Hauptnormale von c normal im Punkt $y = -d\,Q : P$ und besitzt gegen die Schmiegebene die Böschung $-Q : P$. Außerdem sind alle d-Geraden im *Schraubtangentenkomplex* $K(d)$ enthalten; die Kongruenz aller eigentlichen d-Geraden besteht daher aus allen *Komplex-*

kegeln von K(d), deren Spitzen auf q liegen[4]. Läßt man d variieren, so erhält man einen *Strahlkomplex*, dessen Gleichung in jenem Koordinatensystem (x', y', z') mit der rektifizierenden Kante k als z'-Achse, das aus dem (x, y, z)-System durch die Bewegung (8) entsteht, nach (9) und (53) lautet

$$(54)\qquad (P p_4' - Q p_6')(p_1'^2 + p_2'^2) - p_3' p_6'(p_1' + p_4') = 0 .$$

In der zu einem festen Wert von d gehörigen Kongruenz von eigentlichen d-Geraden liegen auch *isotrope Geraden*, die man unter Beachtung von (51) erhält zu

$$(55)\qquad \begin{aligned} p_1 : \ldots : p_6 = \frac{Q}{P} : \sqrt{-\frac{Q^2+P^2}{P^2}} : 1 : \frac{Q}{P}(d-\mu_2) : \\ : \sqrt{-\frac{P^2}{Q^2+P^2}}\left[-d\frac{Q^2}{P^2} - d + \frac{Q}{P}\mu_1 + \frac{Q^2}{P^2}\mu_2\right] : \\ : \left(d - \frac{Q}{P}\mu_1\right). \end{aligned}$$

Diese Geraden besitzen die von d unabhängige Richtung

$$(56)\qquad Q : i\sqrt{P^2+Q^2} : P$$

und bilden die isotropen Parallelstrahlbüschel in den beiden isotropen Ebenen

$$(57)\qquad Qx + i\sqrt{P^2+Q^2}\,y + Pz + id\frac{Q}{P}\sqrt{P^2+Q^2} = 0$$

durch die Achse q. Läßt man d variieren, so erhält man zwei Parallelstrahlbündel mit den festen Richtungen (56) als Gesamtheit aller isotropen eigentlichen d-Geraden.

Satz 7: *Sämtliche eigentlichen d-Geraden einer nicht ebenen Böschungslinie erfüllen einen kubischen Strahlkomplex, in dem zwei Parallelstrahlbündel isotroper Geraden enthalten sind. Die zu einem festen Wert von d gehörigen eigentlichen d-Geraden bilden jene Strahlkongruenz, die aus allen Komplexkegeln des Schraubtangentenkomplexes K(d) besteht, deren Spitzen auf einer gewissen, die Hauptnormale senkrecht schneidenden Geraden liegen; in dieser Kongruenz liegen zwei Parallelbüschel isotroper eigentlicher d-Geraden.*

5. Wir betrachten nun die *Bertrandkurven* und setzen

$$(58)\qquad \mu = Q : P\,, \quad \lambda = \frac{1}{2d}\left(-a \pm \sqrt{a^2 - 4d(d-b)}\right).$$

Nach (50) erfüllen alle eigentlichen d-Geraden zu jedem der beiden Werte von λ eine *Strahlfläche* mit der zur Hauptnormale parallelen Richtebene

$$(59)\qquad p_1 - \lambda p_3 = 0\,,$$

deren Gleichung nach Elimination des Parameters μ lautet[5])

$$(60)\qquad d\lambda^2 z^2 + (b-d)x^2 - b\lambda xz - (2d-b)^2\lambda y + (2d-b)^2(b-d) = 0\,.$$

---

[4]) Für $d = 0$ findet sich diese Tatsache schon bei Kruppa [3].

[5]) Diese Gleichung findet sich im Sonderfall $d = 0$ bereits bei Kruppa ([3] S. 330), sie ist jedoch leider durch Druckfehler entstellt und sollte in der dortigen Bezeichnung richtig lauten: $2C_1 xz - C_0 x^2 - 4C_1^2(C_0 + y) = 0$.

Das ist ein *hyperbolisches Paraboloid* mit der Hauptnormale als Achse und dem Scheitel $y = (b - d) : \lambda$, $x = z = 0$, das durch eine Ähnlichkeit zwischen der Scheitelerzeugenden $x : z = d\lambda : (b - d)$, $y = (b - d) : \lambda$ und der Ferngeraden der Richtebene (59) erzeugt wird. Läßt man d variieren, so entsteht eine *Strahlkongruenz*, deren Gleichungen nach (50) und (58) lauten

$$p_4(p_1^2 + p_3^2) - (a p_3 + b p_1)(p_1^2 + p_3^2) = \\ = p_6(p_1^2 + p_3^2) - (a p_1 - b p_3)(p_1^2 + p_3^2) + p_3(b p_1^2 + 2a p_1 p_3 - b p_3^2) = 0 . \tag{61}$$

Die *isotropen eigentlichen d-Geraden* erfüllen nach (51) mit $\lambda = \mu = P : Q$ die *Strahlfläche*

$$p_1 : \ldots : p_6 = \mu : \sqrt{-\mu^2 - 1} : 1 : \mu(d - b) : \\ : \sqrt{\frac{-1}{\mu^2 + 1}}\,[(b - d)\mu^2 + a\mu - d] : (d - a\mu) . \tag{62}$$

Sie ist im Gewindebüschel

$$p_4 + (b - d)p_1 + \Lambda(p_6 + a p_1 - d p_3) = 0 \tag{63}$$

enthalten. Die Parameterwerte $\Lambda_{1,2}$ der beiden singulären Gewinde dieses Büschels berechnen sich aus

$$(d - b) - a\Lambda + d\Lambda^2 = 0 \tag{64}$$

und die Trägergeraden q_1, q_2 der beiden Gebüsche sind

$$1 : 0 : \Lambda_{1,2} : (b - d + a\Lambda_{1,2}) : 0 : -d\Lambda_{1,2} , \tag{65}$$

also zwei die Hauptnormale normal schneidende Geraden. Die Gleichung der Fläche (62) lautet[6])

$$x^2 y^2 + [(y - a)z - b x]^2 + y^2(y - a)^2 + d^2[x^2 + z^2 + 2y(y - a)] + \\ + 2d[-a x z - b x^2 - b y(y - a)] + d^2(b - d)^2 = 0 . \tag{66}$$

Diese *Strahlfläche 4. Grades* besteht aus allen Geraden des Netzes mit den Brennlinien q_1 und q_2, die den absoluten Kegelschnitt treffen; sie ist daher von *VII*. Art ([5], S. 262). Läßt man d variieren, so erhält man nach (62) und (63) die *Kongruenz*

$$a p_1^2 - b p_1 p_3 + p_1 p_6 - p_3 p_4 = p_1^2 + p_2^2 + p_3^2 = 0 \tag{67}$$

von isotropen eigentlichen d-Geraden.

Satz 8: *Die eigentlichen d-Geraden einer Bertrandkurve erfüllen, soweit für sie* $p_1 \neq 0$ *und* $p_3 \neq 0$ *gilt, eine* (9,9)-*Kongruenz; zu jedem Wert von d* $\neq 0$ *gehören zwei Strahlscharen zweiten Grades, die für d* $= 0$ *zusammenfallen. Die isotropen eigentlichen d-Geraden erfüllen eine* (4,4)-*Kongruenz; zu jedem Wert von d gehört eine Strahlfläche 4. Grades VII. Art.*

Dies gilt in analoger Weise auch für die *Kurven konstanter Torsion* ($a = 0$) und *windschiefen Kreise* ($b = 0$). Nur bei den letzteren tritt bei den 0-Geraden eine Spezialisierung ein, weil aus $b = d = 0$ nach (49) $\lambda = 0$ und damit nach (10) auch $P = 0$ folgt, also nicht mehr der bisher vorausgesetzte allgemeine Fall mit $P \neq 0$ vorliegt.

---

[6]) Speziell für den Fall $d = 0$ findet sich diese Gleichung schon bei Kruppa ([3] S. 330).

6. Die Formeln (50) versagen, falls $p_2 = Q\lambda + P = 0$ gilt, was nach Satz 4 nur bei zerfallenden Cesàrokurven möglich ist. Nach (23) und (14) gilt dann

$$d = \frac{Q}{P^2 + Q^2}(Pa + Qb) \tag{68}$$

und nach (6) für $p_1 \neq 0, p_3 \neq 0 (P \neq 0)$ unter Verwendung von $\lambda = -P : Q$

$$p_1 : \ldots : p_6 = \lambda : 0 : -1 : \frac{\lambda b - a}{\lambda^2 + 1} : p_5 : \lambda \frac{\lambda b - a}{\lambda^2 + 1}. \tag{69}$$

Ist c speziell eine *Böschungslinie*, so ist λ konstant und die eigentlichen d-Geraden haben die Richtung des *Darbouxvektors*

$$\mathfrak{d} = \tau\,\mathfrak{t} + \varkappa\,\mathfrak{b} = \frac{\tau}{P}(P\mathfrak{t} - Q\mathfrak{b}) \tag{70}$$

und sind daher normal zur Bezugsebene der Böschungslinie. Jede solche Gerade erzeugt bei Bewegung des Dreibeins längs c einen *Zylinder*. Ist c eine *Bertrandkurve*, so folgt aus (68) und (69)

$$d(1 + \lambda^2) + a\lambda - b = 0\,, \quad y = d\lambda + a\,. \tag{71}$$

Die zu einem festen Wert von d gehörigen eigentlichen d-Geraden erfüllen je ein Parallelstrahlbüschel in zwei Ebenen normal zur Hauptnormale, für $d = 0$ ein Parallelstrahlbüschel. Läßt man d variieren, so erhält man ein parabolisches Netz um die Ferngerade der rektifizierenden Ebene. Analoges gilt auch für windschiefe Kreise und Kurven konstanter Torsion.

Ist außerdem $P = 0$, so ist die Böschungslinie eine *ebene Kurve*. Für sie ergeben sich für $p_1 = p_2 = 0$ die Normalen ihrer Ebene; ist c eine *Bertrandkurve*, so liefert $p_1 = p_2 = 0$ ein zur Binormale paralleles Strahlbüschel des obigen parabolischen Netzes, wobei für alle Strahlen $d = b$ gilt[7]). Für $p_2 = p_3 = 0$ und damit $Q = 0$ lautet die Cesàrokurve

$$\varkappa(A\varkappa + C\tau) = 0\,, \tag{72}$$

d. h. es ist $1 : a = 0$ und $1 : b = 0$, und zwar so, daß $Pa = A\,(\neq 0)$, $Pb = C\,(\neq 0)$. Aus (4) folgt

$$A : C = -d : y\,. \tag{73}$$

Alle zur Tangente einer Böschungslinie parallelen Geraden sind daher eigentliche d-Geraden, wobei zu festem d jene ∞^1 Parallelen gehören, die die Normalebene in der zur Binormale parallelen Geraden $y = -dC : A$ durchsetzen[8]). Bei Bertrandkurven gibt es keine zur Tangente parallele eigentliche d-Gerade.

Satz 9: *Die zur rektifizierenden Ebene parallelen eigentlichen d-Geraden sind bei Böschungslinien (außer den Parallelen zur rektifizierenden Kante) alle zur Tangente parallelen Geraden, wobei die zum gleichen Wert von d gehörigen eine zur rektifizierenden Ebene parallele Ebene erfüllen. Bei den Bertrandkurven gibt es ein parabolisches Netz um die Ferngerade der rektifizierenden Ebene, zu jedem Wert von d(≠0) zwei Parallelstrahlbüschel in zur Hauptnormale normalen Ebenen.*

---

[7]) Darin steckt folgender bekannter Sonderfall: *Die Binormalenfläche einer Kurve konstanter Torsion τ hat den konstanten Drall* $1 : \tau$.

[8]) Diese Tatsache habe ich bereits in [4] abgeleitet.

Nun fehlt noch die Diskussion im Falle $P = 0$ und $p_2 \neq 0$. Es sei zunächst $P = p_1 = 0$, $p_3 \neq 0$. Aus (32) und (14) folgt

$$d = b \tag{74}$$

und aus (33)

$$p_1 : \ldots : p_6 = 0 : \mu : 1 : a(\mu^2 + 1) : b\mu : -b\mu^2 . \tag{75}$$

Ist c eine *Böschungslinie*, so ist sie wegen $P = 0$ *eben* und (75) stellt alle ∞^3 zur Normalebene von c parallelen Geraden dar. Zu jedem festen Wert von $d = b$ gehört die *Strahlkongruenz*

$$p_1 = p_5 - b p_2 = 0 , \tag{76}$$

die als Schnitt eines Gebüsches mit dem Normalengewinde einer Schraubung um die Hauptnormale vom Parameter $-b$ entsteht; diese Kongruenz ist ein *parabolisches Netz* um die Ferngerade der Normalebene. Ist c eine *Bertrandkurve*, so erfüllen nach (75) alle eigentlichen d-Geraden ein *hyperbolisches Paraboloid*

$$b x z + b^2 y - a(x^2 + b^2) = 0 , \tag{77}$$

in das die Strahlfläche Φ (34) in diesem Falle übergeht. Sämtliche eigentlichen d-Geraden liefern Strahlflächen vom *gleichen konstanten Drall* $d = b$. Ist die Bertrandkurve speziell ein *windschiefer Kreis* ($b = 0$), so bilden die zugehörigen 0-Geraden die Tangenten der *Parabel*

$$z^2 + 4ay - 6a^2 = 0 , \quad x = 0 \tag{78}$$

der Normalebene, deren Scheitel der Krümmungsmittelpunkt von c ist und deren Brennpunkt auf c liegt [1], [3].

Es sei nun $P = p_3 = 0$, $p_1 \neq 0$. Aus (38) und (14) folgt $d = 0$ und aus (39)

$$p_1 : \ldots : p_6 = \mu^2 : \mu : 0 : b : -\mu b : a . \tag{79}$$

Bei einer *ebenen Kurve* sind das alle zur Schmiegebene parallelen Geraden, bei einer *Bertrandkurve* das *hyperbolische Paraboloid*[9])

$$b x z + a z^2 + b^2 y = 0 . \tag{80}$$

Dieses fällt bei einer Kurve konstanter Torsion ($a = 0$) mit dem hyperbolischen Paraboloid (60) zusammen. Für einen *windschiefen Kreis* ($b = 0$) erfüllen die 0-Geraden (79) die Tangenten der *Parabel*

$$x^2 - 4ay = 0 , \quad z = 0 \tag{81}$$

in der Schmiegebene mit dem Scheitel auf c und dem Brennpunkt in der Krümmungsmitte von c [1], [3].

Der Fall $P = p_1 = p_3 = 0$ liefert nichts Neues.

Satz 10: *Die eigentlichen d-Geraden einer ebenen Kurve erfüllen das Gebüsch der Ferngeraden der Normalebene; zu jedem Wert von d gehört ein parabolisches*

---

[9]) Diese Gleichung findet sich schon — bis auf einen Druckfehler — bei Kruppa ([3] S. 332). Sie müßte in der dortigen Bezeichnung lauten: $E_0 x z + 2D_0 z^2 + E_0^2 y = 0$.

Netz; außerdem sind alle Parallelen zur Ebene der Kurve 0-Geraden. Bei den Bertrandkurven existieren außer den in Satz 8 genannten eigentlichen d-Geraden noch eine zur Normalebene parallele Strahlschar eines hyperbolischen Paraboloids, deren Strahlen auf Flächen vom gleichen Drall b führen, und eine zur Schmiegebene parallele Strahlschar eines hyperbolischen Paraboloids von 0-Geraden. Speziell für Kurven konstanter Torsion arten die beiden Strahlscharen, die für windschiefe Kreise zusammenfallen, in die Tangenten zweier Parabeln aus.

7. Abschließend kann nun die Frage beantwortet werden, ob bei jeder Cesàrokurve eigentliche d-Geraden existieren, d. h. ob die Bedingung (3) auch *hinreichend* ist. Aus den vorhergehenden Sätzen folgt, daß bei jeder Cesàrokurve eigentliche d-Geraden existieren, bei den zerfallenden sogar zu jedem beliebig vorgegebenen Wert von d.

Gibt man jedoch bei einer nicht zerfallenden Cesàrokurve ($\Delta \neq 0$) den Wert von d vor, so existieren in einigen Sonderfällen *keine eigentlichen d-Geraden*, wie die Diskussion der Sätze 3 und 6 in Verbindung mit (6), (20) und (21) lehrt. Ist nämlich außer $\Delta \neq 0$ noch $P \neq 0$ (allgemeiner Fall), so darf wegen $\varrho P = p_1 p_3$ die Gleichung (20) nicht als einzige Lösung die Werte $\lambda = 0$ oder $\lambda = \infty$ besitzen; außerdem darf in diesem Fall wegen Satz 3 der Wert $\lambda = -P : Q$ nicht als einzige Lösung neben $\lambda = 0$ und $\lambda = \infty$ auftreten. (20) darf also nicht die Bauart haben $(\lambda - \lambda_1)(\lambda - \lambda_2)(\lambda - \lambda_3) = 0$, wobei $\lambda_i (i = 1, 2, 3)$ nur eine der Zahlen $0, \infty, -P : Q$ ist. Analoges gilt für $d = 0$ für Gleichung (21). Das ergibt:

Satz 11: *Eine nicht zerfallende Cesàrokurve mit $P \neq 0$ besitzt dann keine eigentlichen d-Geraden, wenn für $d \neq 0$ gilt*

$$(I) \quad d = AQ : (3P^2 + Q^2) = BQ^3 : (P^4 + Q^4) = CQ^2 : (PQ^2 - 3P^3) \quad \textit{oder}$$

$$(II) \quad d = Q(AP^2 + CPQ + BQ^2) : (P^2 + Q^2)^2 = Q(2AP + CQ) : \\ : 3P(P^2 + Q^2) = B : Q \quad \textit{oder}$$

$$(III) \quad d = Q(AP^2 + CPQ + BQ^2) : (P^2 + Q^2)^2 = B : Q = C : P \quad \textit{oder}$$

$$(IV) \quad d = B : Q = C : P = A : Q,$$

bzw. für $d = 0$

$$(I') \quad AP^2 - BQ^2 = 0, \quad 2AP + CQ = 0 \quad \textit{oder}$$

$$(II') \quad A : C = -Q : P, \quad B = 0 \quad \textit{oder}$$

$$(III') \quad B : C = -P : Q, \quad A = 0 \quad \textit{oder}$$

$$(IV') \quad B = C = 0 \quad \textit{oder}$$

$$(V') \quad A = B = 0 \quad \textit{oder}$$

$$(VI') \quad A = C = 0.$$

Ist nun $P = 0$, so gilt nach Satz 6 entweder $d = B : Q$ oder $d = A : Q(1 + \mu^2)$. Nur für $d = 0$, $A \neq 0$, $B \neq 0$ ist keine der beiden Gleichungen erfüllbar.

Satz 12: *Eine nicht zerfallende Cesàrokurve mit $P = 0$ und $A \neq 0$, $B \neq 0$ besitzt keine eigentlichen 0-Geraden (VII').*

Sieht man von diesen Sonderfällen[10]) ab, so gilt:

[10]) Salkowski hat in [3] für $d = 0$ die Bedingungen II'—VII' angegeben, den Fall I' aber übersehen. Ein Beispiel dieser Art: $(\varkappa - \tau)^2 = \varkappa + \tau$.

Satz 13: *Bei jeder Cesàrokurve existieren eigentliche d-Geraden. Schreibt man den Wert von d vor, so existieren bei zerfallenden Cesàrokurven stets eigentliche d-Geraden, bei nicht zerfallenden nur dann, wenn kein in Satz 11 und 12 genannter Sonderfall vorliegt.*

Literatur

[1] Cesàro, E.: Vorlesungen über natürliche Geometrie. 2. Aufl. 189ff. Leipzig 1926. — [2] Salkowski, E.: Die Cesàroschen Kurven. S.-B. Akad. Wiss. München. 1911, 523—537. — [3] Kruppa, E.: Strahlflächen als Verallgemeinerung der Cesàro-Kurven. Mh. Math. 52, 323—336 (1948). — [4] Brauner, H.: Über Strahlflächen von konstantem Drall. Mh. Math. 63, 101—111 (1959). — [5] Müller, E., u. J. Krames: Konstruktive Behandlung der Regelflächen. Vorlesungen über Darstellende Geometrie. III. Wien 1931. — [6] Andrade, J.: Sur les droites de contact des courbes gauches et sur une famille de courbes gauches. C. R. Acad. Sci. (Paris) 122, 1110—1113 (1896).

(Eingegangen am 22. Januar 1959)

Voss, K.
Math. Annalen 138, 42—54 (1959)

Über geschlossene Weingartensche Flächen

Von
K. Voss in München*)

1. Auf einer Fläche F im dreidimensionalen euklidischen Raum, die mindestens dreimal stetig differenzierbar ist, seien k_1, k_2 die beiden Hauptkrümmungen, $H = \frac{1}{2}(k_1 + k_2)$ die mittlere Krümmung und $K = k_1 k_2$ die Gaußsche Krümmung. F heißt eine *Weingartensche Fläche* — kurz *W-Fläche* — wenn die beiden Funktionen H und K abhängig sind, genauer gesagt, wenn die Funktionaldeterminante von H, K nach lokalen Flächenparametern u, v identisch verschwindet:

$$\frac{\partial(H,K)}{\partial(u,v)} \equiv 0 \,. \tag{1}$$

Im folgenden wird besonders die Frage untersucht, welche *geschlossenen* W-Flächen es gibt. Bekannte Beispiele solcher Flächen sind die Kugel, die übrigen geschlossenen Rotationsflächen (vom Geschlecht 0 oder 1) und die geschlossenen Röhrenflächen (vom Geschlecht 1). Daß diese Flächen zu den W-Flächen gehören, ist leicht zu sehen: Bei Rotationsflächen hängen H und K nur von einem Parameter ab; bei Röhrenflächen ist eine Hauptkrümmung konstant, also läßt sich z. B. K durch H ausdrücken. Durch Zusammensetzen geeigneter Stücke von Flächen der genannten Arten kann man weitere geschlossene W-Flächen konstruieren, darunter auch solche beliebigen Geschlechts, die noch beliebig oft differenzierbar sind. Beschränkt man sich aber auf *analytische* Flächen — und im folgenden wird dies geschehen — so sind die oben genannten Beispiele die einzigen, die ich kenne.

In dieser Arbeit wird bewiesen:

Satz 1: *Die einzigen analytischen geschlossenen W-Flächen vom topologischen Typus der Kugel sind Rotationsflächen*[1]).

2. Zum Beweis des Satzes 1 verwende ich ein Verfahren, das von S. Cohn-Vossen[2]) in die Differentialgeometrie eingeführt worden ist und im Zusammenhang mit W-Flächen schon von H. Hopf verwendet wurde. Hierbei

*) Die vorliegende Arbeit entstand während eines Aufenthalts am Mathematischen Institut der Universität Freiburg i. Br. im Rahmen eines Stipendiums der Deutschen Forschungsgemeinschaft. Mein besonderer Dank gilt Herrn Prof. W. Süss für seine freundliche Unterstützung und Anregung. Über die Ergebnisse der Arbeit wurde bereits am Internationalen Mathematikertreffen in Wien 1956 berichtet.

[1]) Hierbei sind beliebige Selbstdurchdringungen zugelassen: vgl. die Definition der geschlossenen Fläche im Raum am Schluß von Nr. 6.

[2]) Literaturangaben am Schluß der Arbeit.

handelt es sich einerseits um die Untersuchung der lokalen Eigenschaften eines Feldes tangentialer Linienelemente, das mit dem differentialgeometrischen Problem verbunden ist; andererseits muß ein Feld auf der ganzen Fläche auf Grund der globalen topologischen Gestalt gewissen bekannten Bedingungen genügen.

Ich betrachte ein Feld *orthogonaler Linienelementpaare*. Ist ein solches Feld in der Umgebung einer isolierten Singularität o definiert und (mit Ausnahme von o) stetig, so wird der *Index* j der Singularität o folgendermaßen definiert: Man bedecke die Umgebung von o mit einem ausnahmslos stetigen Feld von Bezugsrichtungen; bei einmaliger Umlaufung von o kehrt das von den beiden Linienelementen gebildete orthogonale Kreuz in sich zurück, erfährt also eine Winkeländerung $\delta\alpha = k\frac{\pi}{2}$ gegenüber der Bezugsrichtung mit einer ganzen Zahl k. Der Index ist dann definiert als $j = \frac{1}{2\pi}\,\delta\alpha = \frac{1}{4}\,k$. Er ist unabhängig von den bei der Definition auftretenden Willkürlichkeiten.

Mit jeder Fläche im Raum ist in bekannter Weise das Feld der Krümmungsrichtungen verbunden. Singularitäten sind die Nabelpunkte, d. h. die Punkte mit $H^2 - K = 0$. In diesem Fall ist k stets gerade, also j eine halbganze Zahl, denn bei Umlaufung einer isolierten Singularität kehrt eine — etwa die zur größeren Hauptkrümmung gehörige — Krümmungsrichtung in sich zurück.

Ist auf einer geschlossenen orientierbaren Fläche vom Geschlecht g ein Tangentialfeld der betrachteten Art mit Ausnahme isolierter Singularitäten definiert, so ist nach dem Satz von Poincaré

(2) $$\Sigma j = 2(1-g)\,.$$

Insbesondere gilt für jedes Feld auf der Kugel

$$\Sigma j = 2\,.$$

Zum Beispiel besitzt das Krümmungsrichtungsfeld auf dem allgemeinen Ellipsoid 4 Singularitäten mit $j = \frac{1}{2}$, dasjenige auf dem Rotationsellipsoid 2 Singularitäten mit $j = 1$.

Die lokalen Eigenschaften der Krümmungslinien auf W-Flächen sollen nun genauer untersucht werden.

3. Hierzu knüpfe ich an Ergebnisse von H. Hopf an. Es sei ein analytisches Flächenstück vorgelegt, welches einen Nabelpunkt o enthält, aber nicht nur aus Nabelpunkten besteht. Wir betrachten speziell solche Flächenstücke, bei denen der Limes

(3) $$\lim \left| \frac{H - H_0}{\sqrt{H^2 - K}} \right| = \lambda$$

bei Annäherung an o existiert (wobei zur Bildung des Limes nur Folgen von Nicht-Nabelpunkten zugelassen sind). Die Klasse dieser Flächen umfaßt alle W-Flächen, wie später gezeigt wird. Ich nenne — in Anlehnung an die Ausdrucksweise bei Differentialgleichungen — den Nabelpunkt o *elliptisch*, *parabolisch* bzw. *hyperbolisch*, je nachdem ob $\lambda < 1$, $\lambda = 1$ oder $\lambda > 1$ ist. Von H. Hopf wurde gezeigt:

a) Die Nabelpunkte vom elliptischen Typus sind isoliert und haben negativen Index.

b) Die Nabelpunkte vom hyperbolischen Typus sind isoliert und haben den Index 1.

Unter der stärkeren Voraussetzung, daß F ein W-Flächenstück ist, läßt sich nun auch über die Nabelpunkte vom parabolischen Typus eine Aussage gewinnen, und das Hopfsche Ergebnis im hyperbolischen Fall läßt sich verschärfen. Ich werde beweisen:

Satz 2: *F sei ein analytisches W-Flächenstück, welches nicht nur aus Nabelpunkten besteht, und o ein Nabelpunkt vom parabolischen Typus auf F. Dann läuft durch o eine reguläre Kurve Γ, gegeben durch eine Gleichung $t(u, v) = 0$, $(t_u, t_v) \neq (0, 0)$, welche ganz aus Nabelpunkten besteht und alle Nabelpunkte von F in der Umgebung von o enthält. Das Paar der Krümmungsrichtungen verhält sich bei Annäherung an Γ stetig, und zwar wird im Limes eine Richtung tangential, die andere orthogonal zu Γ.*

Satz 3: *o sei ein Nabelpunkt vom hyperbolischen Typus auf einem analytischen W-Flächenstück F. Dann ist F rotationssymmetrisch bezüglich der Flächennormalen in o als Rotationsachse.*

4. Beweis des Satzes 1: Es sei F eine geschlossene W-Fläche vom Geschlecht 0. Wir können annehmen, daß F nicht nur aus Nabelpunkten besteht. Wir erweitern das Netz der Krümmungsrichtungen zu einem Netz mit isolierten Singularitäten, indem wir in Nabelpunkten vom parabolischen Typ — falls solche vorhanden sind — als Netzrichtungen die Tangential- und Orthogonalrichtung zu der Nabellinie des Satzes 2 definieren. Da die Indexsumme positiv ist, besitzt das erweiterte Feld eine Singularität o vom hyperbolischen Typ. Somit ist F nach Satz 3 in der Umgebung von o rotationssymmetrisch; die analytische Fläche F hat also mit der gedrehten Fläche ein Flächenstück gemeinsam, also ist F mit der gedrehten Fläche identisch, also F eine Rotationsfläche.

5. Falls $W(\xi, \eta)$ eine reelle Funktion von zwei Variablen ist, die etwa als symmetrisch, analytisch und nicht identisch verschwindend angenommen werde, so kann man die Klasse aller W-Flächen betrachten, zwischen deren Hauptkrümmungen die Relation

$$(W) \qquad W(k_1, k_2) = 0$$

besteht.

Durch die Gleichung (W) wird in der k_1, k_2-Ebene ein Kurvengebilde von lokal-algebraischem Charakter definiert; den Nabelpunkten entsprechen Punkte mit $k_1 = k_2$, also Lösungen der Gleichung $W(k, k) = 0$, und zu jedem solchen Punkt gehören eine oder mehrere Zahlen $\varkappa = \dfrac{d k_2}{d k_1}$ für die einmündenden reellen Kurvenzweige.

Die Sätze 2 und 3 gestatten auch noch gewisse Aussagen über Flächen höheren Geschlechts, wenn man nur spezielle Relationen (W) betrachtet; z. B. gilt:

Satz 4: *Die Relation* (W) *sei so beschaffen, daß alle oben definierten Zahlen* $\varkappa \geqq 0$ *sind. Dann gibt es — abgesehen von eventuellen Rotationsflächen vom Geschlecht* 0 *— keine weiteren analytischen W-Flächen vom Geschlecht* $g \neq 1$ *mit der Relation* (W).

Beispiel: Die lineare Relation

$$(L) \qquad k_2 = \varkappa k_1 + \mu\,, \quad \varkappa, \mu \text{ const}\,, \quad \varkappa \geqq 0\,,$$

bzw. die zugehörige symmetrische Relation

$$(k_2 - \varkappa k_1 - \mu)\,(k_1 - \varkappa k_2 - \mu) = 0$$

hat die im Satz 4 verlangte Eigenschaft. Daher sind die *einzigen analytischen Flächen vom Geschlecht* $g \neq 1$ *mit einer Relation* (L) *Rotationsflächen*. Für $\varkappa^{\pm 1} = 3, 5, 7, \ldots$ und nur für diese Werte von $\varkappa$ gibt es tatsächlich derartige geschlossene Rotationsflächen (vgl. H. Hopf).

Beweis des Satzes 4: Da die Indexsumme für $g \neq 1$ von Null verschieden ist, gibt es Nabelpunkte, und nach Satz 2 sogar nicht-parabolische Nabelpunkte. Diese müssen wegen der speziellen Art der Relation (W) vom hyperbolischen Typ sein, denn in jedem Nabelpunkt ist $\lambda = |(1+\varkappa)/(1-\varkappa)| \geqq 1$. Also ist die Fläche nach Satz 3 eine Rotationsfläche mit $g = 0$.

Für $g = 1$ läßt sich der Schluß nicht anwenden, da es auf einer solchen Fläche keine Nabelpunkte zu geben braucht. Aus den Sätzen 2 und 3, kombiniert mit der Indexsummenformel (2), folgt aber allgemein:

Satz 5: *Auf einer analytischen W-Fläche vom Geschlecht* $g = 1$ *kann es höchstens Nabelpunkte vom parabolischen Typus geben. Auf einer Fläche mit* $g \geqq 2$ *gibt es elliptische, aber keine hyperbolischen Nabelpunkte.*

Im folgenden werden zunächst die benötigten Formeln der Flächentheorie zusammengestellt. Dann wird die Abweichung zweier sich berührender Flächen und ihre Wirkung auf die Krümmungsgrößen untersucht. Danach werden die Sätze 2 und 3 bewiesen: dabei beruht der erste Schritt auf einem algebraischen Lemma von H. Hopf; später werden außer einem funktionentheoretischen Satz über abhängige Funktionen nur elementare Eigenschaften homogener Formen benutzt.

Definition und Darstellung der Flächen

6. Im dreidimensionalen euklidischen Raum sei ein cartesisches Koordinatensystem eingeführt. Ein *Flächenstück* ist gegeben durch den Ortsvektor

$$\mathfrak{x} = \{x, y, z\}\,,$$

wobei x, y, z reelle Funktionen zweier Parameter u, v in einer Umgebung des Nullpunktes der u, v-Ebene sind. Die linear unabhängigen Vektoren $\mathfrak{x}_u, \mathfrak{x}_v$ spannen die Tangentialebene auf; $\mathfrak{u}$ sei der Einheitsvektor in Richtung $-(\mathfrak{x}_u \times \mathfrak{x}_v)$. In den Ableitungsgleichungen

$$\mathfrak{u}_u = \alpha \mathfrak{x}_u + \gamma \mathfrak{x}_v$$
$$\mathfrak{u}_v = \beta \mathfrak{x}_u + \delta \mathfrak{x}_v$$

erfüllt der Abbildungstensor

(4) $$\begin{pmatrix}\alpha & \beta\\ \gamma & \delta\end{pmatrix}$$

die Symmetriebedingung $\mathfrak{u}_u\mathfrak{x}_v = \alpha F + \gamma G = \mathfrak{u}_v\mathfrak{x}_u = \beta E + \delta F$ bezüglich des metrischen Fundamentaltensors $E = \mathfrak{x}_u^2$, $F = \mathfrak{x}_u\mathfrak{x}_v$, $G = \mathfrak{x}_v^2$, d. h. es gilt

(5) $$-\beta E + (\alpha - \delta)F + \gamma G = 0 .$$

Die Eigenwerte des Tensors (4) sind die Hauptkrümmungen, die Eigenvektoren die Krümmungsrichtungen in der u, v-Ebene, also ist

$$2H = \alpha + \delta\,, \quad K = \alpha\delta - \beta\gamma\,,$$

und die Krümmungsrichtungen sind die Nullrichtungen der quadratischen Form

(6) $$-\gamma\, du^2 + (\alpha - \delta)\, du\, dv + \beta\, dv^2 = 0 .$$

Haben die beiden Hauptkrümmungen in einem Nabelpunkt den gemeinsamen Wert c, so gilt in jedem Parametersystem

$$\begin{pmatrix}\alpha & \beta\\ \gamma & \delta\end{pmatrix} = \begin{pmatrix}c & 0\\ 0 & c\end{pmatrix}.$$

Die Codazzischen Gleichungen für den Tensor (4) haben die Form von Kongruenzen

(7) $$\begin{matrix}\alpha_v - \beta_u \equiv 0\\ \gamma_v - \delta_u \equiv 0\end{matrix} \qquad \mathrm{mod}\,(\alpha - \delta, \beta, \gamma)\,.$$

Zur Untersuchung der Umgebung eines Flächenpunktes o läßt man den Nullpunkt des räumlichen Koordinatensystems mit o, die x, y-Ebene mit der Tangentialebene in o zusammenfallen und geht durch eine Parametertransformation von u, v zu x, y als lokalen Flächenparametern über. Dann wird die Fläche durch eine Funktion

$$z = z(x, y) \quad \text{mit} \quad z(0,0) = z_x(0,0) = z_y(0,0) = 0$$

dargestellt. Ich setze

$$z_x = p\,, \quad z_y = q\,, \quad 1 + p^2 + q^2 = W^2\,, \; W > 0\,,$$
$$\frac{p}{W} = P\,, \quad \frac{q}{W} = Q\,.$$

Damit wird

$$\mathfrak{u} = \{P, Q, -W^{-1}\}\,,$$

und die Komponenten des Tensors (4) berechnen sich im x, y-System nach der Formel

$$\begin{pmatrix}\alpha & \beta\\ \gamma & \delta\end{pmatrix} = \begin{pmatrix}P_x & P_y\\ Q_x & Q_y\end{pmatrix}.$$

Eine *geschlossene Fläche* vom Typus der Kugel ist folgendermaßen definiert: Auf einer Parameterkugel sind drei Ortsfunktionen x, y, z gegeben mit der Eigenschaft: falls u, v lokale Parameter auf der Kugel sind, so liefert der zugehörige Ortsvektor $\mathfrak{x}$ ein Flächenstück. $\mathfrak{x}$ definiert also eine Abbildung der

Parameterkugel in den Raum, und zwar eine differenzierbare Abbildung vom Rang 2. Die Abbildung braucht aber keineswegs eineindeutig zu sein.

Eine Fläche heißt (reell) *analytisch*, wenn die auftretenden Funktionen sich durch konvergente Potenzreihen in u, v darstellen lassen.

Berührung analytischer Flächen

7. Sind W, P, Q wie oben definiert und p_1, q_1, p_2, q_2 beliebige Werte der Variablen p, q, so gilt für die Differenzen $P_1 - P_2$, $Q_1 - Q_2$, wie man nachrechnet, eine Darstellung

$$\begin{aligned} P_1 - P_2 &= \varkappa(p_1 - p_2) + \lambda(q_1 - q_2) \\ Q_1 - Q_2 &= \mu(p_1 - p_2) + \nu(q_1 - q_2) \end{aligned} \tag{8}$$

mit

$$\varkappa = \frac{1}{W_1}\left\{1 - P_2\frac{p_1 + p_2}{W_1 + W_2}\right\} \qquad \lambda = \frac{1}{W_1}\left\{\; - P_2\frac{q_1 + q_2}{W_1 + W_2}\right\}$$
$$\mu = \frac{1}{W_1}\left\{\; - Q_2\frac{p_1 + p_2}{W_1 + W_2}\right\} \qquad \nu = \frac{1}{W_1}\left\{1 - Q_2\frac{q_1 + q_2}{W_1 + W_2}\right\}.$$

Nun seien F_1, F_2 zwei analytische Flächenstücke, die sich in o berühren, mit den Darstellungen

$$z = z_1(x, y)\,, \quad z = z_2(x, y)\,.$$

Ich setze

$$z_1 - z_2 = d(x, y)$$

und entwickle d nach homogenen Formen i-ten Grades in x, y: $d = \Sigma d^{(i)}$. Wenn F_1, F_2 nicht identisch sind, beginnt die Entwicklung mit

$$d(x, y) = d^{(k)}(x, y) + \cdots, \qquad k \geqq 2\,, \quad d^{(k)} \not\equiv 0\,.$$

In der Formel (8) ist jetzt $p_1 - p_2 = d_x$, $q_1 - q_2 = d_y$ zu setzen, und wegen $p_i(0, 0) = q_i(0, 0) = 0$ wird $\varkappa(0, 0) = \nu(0, 0) = 1$, $\lambda(0, 0) = \mu(0, 0) = 0$. Daher beginnt die Entwicklung von (8) mit

$$P_1 - P_2 = d_x^{(k)} + \cdots, \qquad Q_1 - Q_2 = d_y^{(k)} + \cdots,$$

und daraus folgt durch Bildung der Funktionalmatrix

$$\begin{pmatrix} \alpha_1 - \alpha_2 & \beta_1 - \beta_2 \\ \gamma_1 - \gamma_2 & \delta_1 - \delta_2 \end{pmatrix} = \begin{pmatrix} d_{xx}^{(k)} & d_{xy}^{(k)} \\ d_{xy}^{(k)} & d_{yy}^{(k)} \end{pmatrix} + \cdots. \tag{9}$$

8. Von jetzt an ist F ein analytisches Flächenstück mit einem Nabelpunkt o und der Krümmung $H_0 = c$ in o. Als *Krümmungskugel* S bezeichne ich die Kugel mit dem Zentrum $\mathfrak{x}_0 - (1/c)\,\mathfrak{u}_0$ und dem Radius $1/|c|$, welche F in o berührt (für $c = 0$ ist S die Tangentialebene in o). S besitzt die Darstellung

$$S: \qquad z = \zeta(x, y) = \frac{c(x^2 + y^2)}{1 + \sqrt{1 - c^2(x^2 + y^2)}}\,.$$

Ich stelle F durch eine Funktion $f(x, y)$ in der Form

$$F: \qquad z = \zeta(x, y) + f(x, y)$$

dar, wobei f die Abweichung von S erfaßt. Wenn F nicht mit S identisch ist,

beginnt f mit einem Anfangsglied

$$f^{(n+2)} = A(x, y) \not\equiv 0\,, \quad n \geqq 1\,.$$

Wendet man (9) auf $F_1 = F$, $F_2 = S$ an, so wird $\alpha_2 = \delta_2 \equiv c$, $\beta_2 = \gamma_2 \equiv 0$, also

$$\begin{pmatrix} \alpha - c & \beta \\ \gamma & \delta - c \end{pmatrix} = \begin{pmatrix} A_{xx} & A_{xy} \\ A_{xy} & A_{yy} \end{pmatrix} + \cdots. \tag{10}$$

Im folgenden werden an Krümmungsgrößen verwendet:

$$H - c = \frac{1}{2}(k_1 - c + k_2 - c) = \frac{1}{2}(\alpha - c + \delta - c)$$

$$H^2 - K = \left(\frac{k_1 - k_2}{2}\right)^2 = \left(\frac{\alpha - \delta}{2}\right)^2 + \beta\gamma$$

$$D = (k_1 - c)(k_2 - c) = (\alpha - c)(\delta - c) - \beta\gamma\,.$$

Aus (10) folgt für die Anfangsglieder

$$H - c = H^{(n)} + \cdots, \qquad H^{(n)} = \frac{1}{2}(A_{xx} + A_{yy})$$

$$H^2 - K = T^{(2n)} + \cdots, \qquad T^{(2n)} = \left(\frac{A_{xx} - A_{yy}}{2}\right)^2 + A_{xy}^2$$

$$D = D^{(2n)} + \cdots \qquad D^{(2n)} = A_{xx}A_{yy} - A_{xy}^2\,.$$

Hierbei ist wegen $A \not\equiv 0$ und $n \geqq 1$ stets $T^{(2n)} \not\equiv 0$, denn die einzigen reellen Funktionen A, für die der Ausdruck $T^{(2n)}$ identisch verschwindet, sind die Polynome $a(x^2 + y^2) + bx + cy + d$ mit Konstanten a, b, c, d.

9. Ich fasse jetzt in der Darstellung

$$F: \qquad z = \zeta + f^{(n+2)} + \cdots + f^{(\nu+2)} + \cdots$$

ein höheres Glied

$$f^{(\nu+2)} = B(x, y)\,, \quad \nu > n\,,$$

ins Auge. F_μ ($\mu \geqq n$) sei die approximierende Fläche

$$F_\mu: \qquad z = \zeta + f^{(n+2)} + \cdots + f^{(\mu+2)}\,.$$

Ich vergleiche F mit der Fläche $F_{\nu-1} = \overset{*}{F}$. Nach (9) gilt

$$\begin{pmatrix} \alpha - \overset{*}{\alpha} & \beta - \overset{*}{\beta} \\ \gamma - \overset{*}{\gamma} & \delta - \overset{*}{\delta} \end{pmatrix} = \begin{pmatrix} B_{xx} & B_{xy} \\ B_{xy} & B_{yy} \end{pmatrix} + \cdots.$$

Führt man auch hier Entwicklungen $\alpha = \Sigma\,\alpha^{(i)}$, $\overset{*}{\alpha} = \Sigma\,\overset{*}{\alpha}{}^{(i)}$ usw. ein, so erhält man für den Anfang der Entwicklung von (10) bis zur Ordnung ν

$$\begin{pmatrix} \sum\limits_{i=n}^{\nu} \alpha^{(i)} & \sum\limits_{i=n}^{\nu} \beta^{(i)} \\ \sum\limits_{i=n}^{\nu} \gamma^{(i)} & \sum\limits_{i=n}^{\nu} \delta^{(i)} \end{pmatrix} = \begin{pmatrix} \sum\limits_{i=n}^{\nu} \overset{*}{\alpha}{}^{(i)} + B_{xx} & \sum\limits_{i=n}^{\nu} \overset{*}{\beta}{}^{(i)} + B_{xy} \\ \sum\limits_{i=n}^{\nu} \overset{*}{\gamma}{}^{(i)} + B_{xy} & \sum\limits_{i=n}^{\nu} \overset{*}{\delta}{}^{(i)} + B_{yy} \end{pmatrix}, \tag{11}$$

wobei die ersten Glieder für $i = n$ in (10) gegeben sind. Aus (11) folgt durch Spurbildung

$$\begin{aligned} &H^{(i)} = \overset{*}{H}{}^{(i)} \quad \text{für} \quad n \leqq i < \nu \\ &H^{(\nu)} = \overset{*}{H}{}^{(\nu)} + \hat{H}^{(\nu)} \quad \text{mit} \quad \hat{H}^{(\nu)} = \frac{1}{2}(B_{xx} + B_{yy}) \end{aligned} \tag{12}$$

und durch Determinantenbildung

$$(13)\quad \begin{aligned} D^{(n+i)} &= \overset{*}{D}{}^{(n+i)} \quad \text{für} \quad n \leqq i < \nu \\ D^{(n+\nu)} &= \overset{*}{D}{}^{(n+\nu)} + \hat{D}^{(n+\nu)} \quad \text{mit} \quad \hat{D}^{(n+\nu)} = A_{yy}B_{xx} - 2A_{xy}B_{xy} + A_{xx}B_{yy}. \end{aligned}$$

10. Später werden noch folgende Hilfssätze gebraucht:

Hilfssatz a: *Wenn $f(x, y) = x^{n+2}g(x, y)$ ist, also F die Krümmungskugel längs des Großkreises $x = 0$ von mindestens $(n + 1)$-ter Ordnung berührt, so ist $H - c = x^n(\ldots)$ und $D = x^{2n+2}(\ldots)$.*

Beweis: Formel (8) für $F_1 = F$ und $F_2 = S$ ergibt wegen $p_1 - p_2 = x^{n+1}(\ldots)$, $q_1 - q_2 = x^{n+2}(\ldots)$ zunächst $P_1 - P_2 = x^{n+1}(\ldots)$. Die Funktion μ enthält den Faktor $p_1 + p_2 = p_1 - p_2 + 2W_2P_2$; wegen $P_2 = cx$ ist also μ ein Vielfaches von x, folglich $Q_1 - Q_2 = x^{n+2}(\ldots)$. Daraus folgt durch Differentiation

$$\begin{pmatrix} \alpha - c & \beta \\ \gamma & \delta - c \end{pmatrix} = \begin{pmatrix} x^n & (\ldots) & x^{n+1}(\ldots) \\ x^{n+1}(\ldots) & x^{n+2}(\ldots) \end{pmatrix}$$

und damit die Behauptung über Spur und Determinante.

Hilfssatz b: *Wenn F analytisch ist, bezogen auf beliebige Parameter u, v, und $H - c = u(\ldots)$, c konstant, und $H^2 - K = u^{2n}g(u, v)$, $n \geqq 1$, $g(0, 0) \neq 0$ gilt*[3]*), so fügen sich für $u = 0$ die v-Richtung und die dazu orthogonale Richtung stetig in das Krümmungsrichtungsfeld ein.*

Beweis: Die Entwicklung von (4) nach Potenzen von u lautet

$$\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} = \begin{pmatrix} c + u^k\alpha_k(v) + \cdots & u^k\beta_k(v) + \cdots \\ u^k\gamma_k(v) + \cdots & c + u^k\delta_k(v) + \cdots \end{pmatrix},$$

wobei $k \geqq 1$ ist und nicht alle vier Funktionen $\alpha_k, \beta_k, \gamma_k, \delta_k$ identisch verschwinden. Man kann annehmen, daß die u-Linien orthogonal zur Kurve $u = 0$ verlaufen; dann ist $F(0, v) = 0$, und aus (5) folgt nach Division durch u^k für $u = 0$:

$$(5_0)\qquad -\beta_k(v)\,E(0, v) + \gamma_k(v)\,G(0, v) = 0\,,$$

wobei $E, G > 0$. Aus (7) folgt $\beta_k(v) \equiv 0$, $\delta_k(v) \equiv 0$, und wegen (5_0) auch $\gamma_k(v) \equiv 0$, also $\alpha_k(v) \not\equiv 0$. Damit bekommt (6) die Gestalt

$$(6')\qquad u^k\,\{u(\ldots)\,du^2 + [\alpha_k(v) + u(\ldots)]\,du\,dv + u(\ldots)\,dv^2\} = 0\,.$$

Wegen $H^2 - K = u^{2k}\left\{\frac{1}{4}\,\alpha_k(v)^2 + u(\ldots)\right\}$ ist $k = n$ und

$$\alpha_k(0) \neq 0\,.$$

Dividiert man (6′) durch u^k, so erhält man eine auch für $u = 0$ stetige Erweiterung des Richtungsfeldes, wobei die Feldrichtungen für $u = 0$ durch $du\,dv = 0$ gegeben sind, q. e. d.

W-Flächen in erster Näherung

11. Jetzt werden Flächen mit Nabelpunkten betrachtet, bei denen der Limes (3) existiert (vgl. die zitierte Arbeit von H. Hopf, S. 242).

---

[3]) Daraus folgt, daß F die Krümmungskugel längs der Nabellinie $u = 0$ von genau $(n + 1)$-ter Ordnung berührt.

Mit den Bezeichnungen von Nr. 8 wird nach Einführung konjugiert komplexer Parameter $w = x + iy$, $\bar{w} = x - iy$

$$H^{(n)} = 2A_{w\bar{w}}\,, \quad T^{(2n)} = |2A_{ww}|^2\,.$$

Die Existenz des Limes (3) ist bei analytischen Flächen gleichbedeutend mit

$$(3') \qquad H^{(n)2} = \lambda^2 T^{(2n)}\,,$$

d. h. gleichbedeutend damit, daß die reelle Form $A(x, y)$ einer Beziehung

$$(3'') \qquad |A_{w\bar{w}}| = \lambda\,|A_{ww}|$$

mit konstantem λ genügt. Nach einem algebraischen Lemma von H. Hopf (angewandt auf $A_{w\bar{w}}$ und A_{ww}) gibt es nur die folgenden Lösungen A, λ der Gleichung (3''): *a)* $\lambda = 0$, *A harmonische Form; b)* $\lambda = 1$, *A Potenz einer Linearform; c)* $\lambda = (m+1)/m$ (m ganz ≥ 1), *A Potenz von $x^2 + y^2$*. Nach einer geeigneten Drehung der x, y-Ebene ist also

(*a*) *im elliptischen Fall* $A = a\,\Re(w^{n+2})$, $\lambda = 0$ $(H^{(n)} = 0)$;

(*b*) *im parabolischen Fall* $A = b\,x^{n+2}$, $\lambda = 1$ $(D^{(2n)} = 0)$;

(*c*) *im hyperbolischen Fall* $A = c(w\bar{w})^{m+1} (m \geq 1)$, $\lambda = (m+1)/m$,

wobei a, b, c reelle Konstanten sind.

In den Fällen (*a*) und (*c*) ist $T^{(2n)}$ eine definite Form, also besitzt $H^2 - K$ eine isolierte Nullstelle, d. h. der Nabelpunkt ist isoliert. Ferner ändert sich der Index nicht, wenn man in der Gleichung (6) zu den Anfangsgliedern übergeht, d. h. zu dem Feld

$$(6_0) \qquad -A_{xy}(dx^2 - dy^2) + (A_{xx} - A_{yy})\,dx\,dy = 2\,\Im(A_{ww}dw^2) = 0\,,$$

welches ebenfalls eine isolierte Singularität hat. Aus (6_0) folgt

$$j = -\frac{1}{4\pi}\,\delta(\arg A_{ww})\,,$$

also $j = -n/2$ im Falle (*a*) und $j = 1$ im Falle (*c*).

Das Ergebnis dieses Abschnitts läßt sich auch folgendermaßen ausdrücken: *Bei einem Nabelpunkt auf einer analytischen Fläche existiert der Limes* (3) *dann und nur dann, wenn die Fläche in ihrem Anfangsglied A entweder mit einer Fläche konstanter mittlerer Krümmung, oder mit einer Röhrenfläche bzw. Torse, oder mit einer rotationssymmetrischen Fläche übereinstimmt.*

12. Zu den Flächen mit Limes (3) gehören insbesondere alle analytischen W-Flächen. Falls nämlich die Funktionaldeterminante (1) identisch verschwindet, so gilt das gleiche auch für die Funktionaldeterminante der beiden Funktionen $(H - c)^2$ und $H^2 - K$ nach x, y, und daraus folgt

$$\frac{\partial\,(H^{(n)2}, T^{(2n)})}{\partial\,(x, y)} = 0\,.$$

Nun gilt folgender Hilfssatz:

Zwei homogene Formen $\Phi(x, y)$, $\Psi(x, y)$ gleichen Grades, deren Funktionaldeterminante verschwindet, sind proportional.

Daraus folgt dann, daß bei W-Flächen (3') gilt, also der Limes (3) existiert.

Zum Beweis des Hilfssatzes führe ich Polarkoordinaten r, ϑ ein und bezeichne die Werte von Φ, Ψ für $r = 1$ mit $\varphi(\vartheta)$ bzw. $\psi(\vartheta)$. Aus dem Verschwinden der Funktionaldeterminante von Φ, Ψ nach r, ϑ folgt

$$\varphi\psi' - \varphi'\psi = 0 ,$$

also ist $\varphi/\psi = \alpha$ konstant in einem Intervall mit $\psi \neq 0$, also verschwindet die Form $\Phi - \alpha\Psi$ identisch.

W-Flächen in höherer Näherung

13. Um die Bedingung (1) in höherer Näherung auszunutzen, gehe ich zu der äquivalenten Gleichung

$$(1') \qquad \Delta = \frac{\partial(H - c, D)}{\partial(x, y)} = 0$$

über und betrachte wie in Nr. 9 die Fläche F *modulo* $F_{\nu-1} = \overset{*}{F}$ für $\nu > n$. Zunächst erhält man aus (1')

$$\Delta^{(2n+\nu-2)} = \sum_{i=n}^{\nu} \frac{\partial(H^{(i)}, D^{(2n+\nu-i)})}{\partial(x, y)} = 0$$

und daraus mit (12) und (13)

$$(14) \qquad \frac{\partial(H^{(n)}, \hat{D}^{(n+\nu)})}{\partial(x, y)} + \frac{\partial(\hat{H}^{(\nu)}, D^{(2n)})}{\partial(x, y)} + \overset{*}{\Delta}{}^{(2n+\nu-2)} = 0 .$$

Betrachtet man $\overset{*}{F}$ als bekannt, so ist (14) eine Differentialgleichung für das nächste Glied $B(x, y)$.

Beweis des Satzes 2

14. Im parabolischen Fall ist nach Nr. 11

$$A = b x^{n+2} , \quad D^{(2n)} = 0 .$$

Durch Ähnlichkeit und eventuelle Spiegelung der Fläche kann man erreichen

$$b = 1/(n + 2)(n + 1) .$$

Das nächste Glied $D^{(2n+1)}$ erhält man aus (13) für $\nu = n + 1$: dort wird erstens $\hat{D}^{(2n+1)} = x^n f_{yy}^{(n+3)}$; zweitens folgt aus dem Hilfssatz a von Nr. 10, angewandt auf $\overset{*}{F} = F_n$: $z = \zeta + A$, daß $\overset{*}{D} = x^{2n+2}(\ldots)$ ist, also $\overset{*}{D}{}^{(2n+1)} = 0$, und damit

$$D^{(2n+1)} = x^n f_{yy}^{(n+3)} .$$

Ich unterscheide zwei Fälle:

1. Fall: *D verschwindet von der Ordnung* $2n + 1$, d. h. $D^{(2n+1)} \not\equiv 0$. Da die beiden reell analytischen Funktionen $H - c$ und D gemäß (1') *abhängig* sind und nicht identisch verschwinden, *gibt es* nach einem lokalen Satz der Funktionentheorie *eine reell analytische Funktion*

$$t(x, y) = t^{(a)} + \cdots , \quad t^{(a)} \not\equiv 0 , \quad a \geq 1 ,$$

und zwei reell analytische Funktionen

$$\varphi(\tau) = \alpha\tau^p + \cdots \quad \text{und} \quad \psi(\tau) = \beta\tau^q + \cdots, \quad \alpha, \beta \neq 0, \quad p, q \geq 1,$$

derart, daß

$$H - c = \varphi(t) \quad \text{und} \quad D = \psi(t)$$

ist[4]).

Da $H^{(n)}$ und $D^{(2n+1)}$ nicht identisch verschwinden, ist

$$ap = n, \quad aq = 2n + 1,$$

also a ein gemeinsamer Teiler von n und $2n + 1$. Diese beiden Zahlen sind aber teilerfremd, also ist $a = 1$. $t(x, y) = 0$ ist die im Satz 2 behauptete Nabellinie.

2. Fall: *D verschwindet mindestens von der Ordnung* $2n + 2$, d. h. $D^{(2n+1)} = 0$. Wegen $f_{yy}^{(n+3)} = 0$ gilt dann

$$f^{(n+3)} = x^{n+2}(\ldots).$$

Ich beweise folgenden

Zusatz zum Satz 2: *Falls D von der Ordnung $\geq 2n + 2$ verschwindet, so ist die Nabellinie Γ ein Kreis senkrecht zur Tangentialebene (bei Flachpunkten: eine Gerade in der Tangentialebene)*[5]).

Ich werde zeigen, daß $f(x, y)$ durch x^{n+2} teilbar ist, daß F also eine Darstellung der Form

$$z = \zeta + x^{n+2}\{b + \cdots\} \tag{15}$$

besitzt. Hierzu beweise ich induktiv, daß alle in Nr. 9 definierten Flächen $F_\nu (\nu = n, n + 1, \ldots)$ eine solche Darstellung haben. Für $\nu = n$ und $\nu = n + 1$ ist dies schon bewiesen. Ich nehme an, für $F_{\nu-1} = \overset{*}{F}$ sei eine Darstellung (15) vorhanden, und betrachte F_ν mit $\nu \geq n + 2$. Für die zu $\overset{*}{F}$ gehörigen Krümmungsgrößen gilt nach dem Hilfssatz a von Nr. 10: $\overset{*}{H} - c = x^n\{\ldots\}$, $\overset{*}{D} = x^{2n+2}\{\ldots\}$, also

$$\overset{*}{\Delta} = x^{3n+1}\{\ldots\}.$$

Ferner ist

$$H^{(n)} = \frac{1}{2}x^n, \quad \hat{D}^{(n+\nu)} = x^n B_{yy}.$$

Damit folgt aus (14)

$$H_x^{(n)} \hat{D}_y^{(n+\nu)} = x^{3n+1}\{\ldots\},$$

also $B_{yyy} = x^{n+2}\{\ldots\}$, und daher

$$B = x^{n+2}\{\ldots\} + a x^\nu y^2 + b x^{\nu+1} y + c x^{\nu+2},$$

[4]) Vgl. W. F. Osgood, S. 164. (Der dortige Beweis enthält eine Unkorrektheit, die aber leicht zu beheben ist.) Daß man im reellen Fall die Funktionen t, φ, ψ ebenfalls reell wählen kann, sieht man so ein: Nach einer geeigneten Transformation des Parameters τ wird $\varphi = \tau^p$, also t^p reell. Ist nun $t(x_0, y_0) \neq 0$, so wird in der Umgebung $\bar{t}(x, y)/t(x, y) = \varepsilon$ eine p-te Einheitswurzel, die aus Stetigkeitsgründen konstant ist. Daher gilt identisch $\bar{t} = \varepsilon t$; mit $\eta^2 = \varepsilon$ wird also ηt reell. — Allgemeinere Sätze globalen Charakters sind neuerdings von K. Stein bewiesen worden (s. insbesondere S. 63—64).

[5]) Hierin sind insbesondere Aussagen über Nabelpunkte auf Röhrenflächen bzw. Torsen enthalten.

wegen $\nu \geqq n + 2$ also

$$B = x^{n+2}(\ldots)\,.$$

Damit ist der Zusatz bewiesen.

Im ersten Fall ist nach Verfügung über eine Konstante

$$t(x, y) = x + \cdots$$

und $H^2 - K = (H - c)^2 - D = t^{2n}\left\{\frac{1}{4} + \cdots\right\}$. Dasselbe gilt auch im zweiten Falle, wenn man $t(x, y) = x$ setzt. Die Behauptung im Satz 2 über das Verhalten der Krümmungsrichtungen bei Annäherung an die Nabelkurve $t(x, y) = 0$ folgt jetzt aus dem Hilfssatz b von Nr. 10 nach der Parametertransformation

$$u = t(x, y)\,, \quad v = y\,.$$

Beweis des Satzes 3

15. Unter Verwendung der Bezeichnungen von Nr. 9 soll gezeigt werden, daß alle Flächen $F_\nu(\nu = n, \ldots)$ Rotationsflächen sind. Nach Nr. 11 ist im hyperbolischen Fall nach geeigneter Normierung

$$A = (w\bar{w})^{m+1}/(n + 2) \qquad (n = 2m)\,,$$

d. h. F_n ist eine Rotationsfläche. Ich betrachte $F_\nu(\nu > n)$ und nehme an, die Behauptung sei für $F_{\nu-1} = \overset{*}{F}$ schon bewiesen. Man erhält in Polarkoordinaten r, ϑ:

$$H^{(n)} = (m + 1)\, r^n\,, \quad D^{(2n)} = (n + 1)\, r^{2n}\,.$$

Ist

$$B = r^{\nu+2} b(\vartheta)$$

die zu untersuchende Form, so wird

$$\hat{H}^{(\nu)} = \frac{1}{2}\, r^\nu h(\vartheta) \quad \text{mit} \quad h = b'' + (\nu + 2)^2 b\,.$$

Weiter ist

$$\begin{aligned}\hat{D}^{(n+\nu)} &= -4\,\{A_{\bar{w}\bar{w}} B_{ww} - 2A_{w\bar{w}} B_{w\bar{w}} + A_{ww} B_{\bar{w}\bar{w}}\}\\ &= r^{n-2}\,\{(n + 2)\, w\bar{w}\, 2 B_{w\bar{w}} - n(w^2 B_{ww} + \bar{w}^2 B_{\bar{w}\bar{w}})\}\,.\end{aligned}$$

Nach Umformung mit der Homogenitätsrelation

$$w^2 B_{ww} + 2w\bar{w} B_{w\bar{w}} + \bar{w}^2 B_{\bar{w}\bar{w}} = (\nu + 2)\,(\nu + 1)\, B$$

erhält man

$$\hat{D}^{(n+\nu)} = r^{n+\nu} d(\vartheta) \quad \text{mit} \quad d = (n + 1)h - n(\nu + 2)\,(\nu + 1) b\,.$$

Da $\overset{*}{F}$ Rotationsfläche ist, verschwindet $\overset{*}{\Delta}$; geht man in (14) noch zu r, ϑ über, so folgt

$$H_r^{(n)} \hat{D}_\vartheta^{(n+\nu)} - D_r^{(2n)} \hat{H}_\vartheta^{(\nu)} = 0$$

oder

$$(m + 1) d' - (n + 1) h' = 0\,.$$

Führt man hier $b(\vartheta)$ ein und ordnet, so erhält man die Bedingung

$$(n+1)b''' = (\nu+2)(\nu-n)b'.$$

Demnach genügt die Funktion $b' = a(\vartheta)$ wegen $\nu > n$ einer Gleichung

$$a'' = \mu a \quad \text{mit} \quad \mu > 0.$$

Da $a(\vartheta)$ periodisch, also eine Funktion auf der Kreislinie ist, gibt es eine Stelle, an der das Maximum angenommen wird; dort ist $a'' \leqq 0$, also auch $a \leqq 0$. Ebenso ergibt die Betrachtung des Minimums $a \geqq 0$. Daraus folgt: $b' = a \equiv 0$, also ist b konstant, q. e. d.

Literatur

Cohn-Vossen, S.: Zwei Sätze über die Starrheit der Eiflächen. Nachr. Ges. Wiss. Göttingen, Math.-Phys. Kl. 1927, 125—134. — Hopf, H.: Über Flächen mit einer Relation zwischen den Hauptkrümmungen. Math. Nachr. 4, 232—249 (1951). — Osgood, W. F.: Lehrbuch der Funktionentheorie. II, 1, 2. Aufl. 1929. — Stein, K.: Analytische Zerlegungen komplexer Räume. Math. Ann. 132, 63—93 (1956).

(Eingegangen am 10. März 1958)

Browder, F. E.
Math. Annalen 138, 55—79 (1959)

Functional Analysis and Partial Differential Equations. I*

By

Felix E. Browder in New Haven, Conn.

Introduction

The general theory of boundary-value problems for linear partial differential operators has seen a flourishing development within the past decade with the successful treatment of the classical types of boundary-value problems for elliptic, hyperbolic, and parabolic differential equations, as well as the general class of equations of propagation. It is the purpose of the present paper, and of its successor, to present a general treatment in unified terms of these various problems and to exhibit explicitly the common methodological basis of the discussion of apparently diverse sorts of problems. In rough terms, this basis consists of the combination of general principles from functional analysis with concrete analytical a-priori inequalities, usually of an "energy" type. We have divided the presentation into two parts. The first, embodied in the present paper, treats exclusively of the relevant methods from functional analysis, while the following paper contains the detailed application of these methods to partial differential operators. The latter, in particular, contains an unified discussion of elliptic boundary-value problems, mixed initial-boundary value problems for equations of propagation, and a simplified treatment of Leray's theory of the Cauchy problem for hyperbolic equations.

We begin with a brief summary of the material presented below. If we are given a differential operator L defined on a family of functions $D(L)$ lying in some function space E and satisfying some linear null boundary conditions on a domain G, we are interested in general in determining the class of functions f in some function space F for which the equation $Lu = f$ has a solution u in $D(L)$, (or in abstract terms, assuming that Lu lies in F for all u in $D(L)$, the range of L). Finite-dimensional linear mappings motivate the introduction of the adjoint operator L^* and the study of the range of L^*. If the finite-dimensional analogue were reliable, the range of L would be the orthogonal complement in some suitable sense of the null space of L^*, i.e. the set of all v in $D(L^*)$ for which $L^* v = 0$. In the infinite dimensional case, this will be true if and only if the range of L is closed in F, and it is this last fact which provides the motivation for the study of operators with closed range given below.

In Sections 1 and 2, we study the inter-relations in a very general linear space setting of various properties of L and L^* centering around the closed

*) This paper was written while the writer held a National Science Foundation Senior Post-Doctoral Fellowship during the academic year 1957—58.

range property. One of our results states that if E and F are Banach spaces, L a closed operator from E to F, then the range of L is closed if and only if the range of L^* is closed. More generally, if E and F are Fréchet spaces, T closed once more, then the range of L is closed if and only if the range of L^* is weak* closed. In Section 1, we derive a number of useful consequences of these and related equivalences, and discuss some other more easily derivable duality principles. Section 2 contains a discussion of the problem of the relation of the closed range and homomorphism properties of L and L^* in the context of the general theory of locally convex linear spaces. In Section 3, we apply the results of Sections 1 and 2 to give a more general treatment of the abstract theory of the existence of solvable boundary-value problems. We give also a more general form of the generalized Friedrichs extension which is applied in the following paper.

A salient general feature of our discussion, and of the whole theory of boundary-value problems, is the significant role played by unbounded (or not necessarily continuous) linear operators. Certainly, in the context of the usual function spaces, differential operators are not continuous operators. One can deal with this difficulty in essentially two ways. The first, as in L. Schwartz's theory of distributions, is to define a new (non-metrizable) topology in which the operators are all continuous. The second, as we shall do below, is to construct a theory of (not necessarily continuous) closed linear operators. Since it is both inconvenient and un-necessary to discard the firm Hilbert or Banach space structures with their numerous useful properties merely to make one's operators fit the Procrustean bed of continuity, the latter approach, which stems primarily from VON NEUMANN, seems to the writer to be the most fruitful and is the basic organizing principle of our discussion.

Section 1: Let E and F be two Banach spaces, E^* and F^* their respective adjoint spaces. We shall denote the pairing between an element e of the space E and an element e^* of its dual space E^* by $\langle e, e^*\rangle$. $E \times F$ will denote the Banach space of pairs $\{(e, f) : e \in E, f \in F\}$, equipped with the usual linear structure and with the norm $\|(e, f)\| = \{\|e\|^2 + \|f\|^2\}^{\frac{1}{2}}$.

Let T be a linear transformation with its domain $D(T)$ a dense subspace of E and its range $R(T)$ contained in F. $N(T)$, the null space of T, consists of all e in $D(T)$ for which $Te = 0$. T is said to be closed if its graph $G(T) = \{(e, f) : e \in D(T), f = Te\}$ is a closed subset of $E \times F$. If T is a closed operator, $N(T)$ is obviously a closed subspace of E.

The adjoint transformation T^* is defined with domain $D(T^*)$ in F^* consisting of all f^* such that $\langle Te, f^*\rangle$ is a bounded linear functional of e in $D(T)$. For each f^* in $D(T^*)$, there exists an unique element e^* of E^* such that $\langle Te, f^*\rangle = \langle e, e^*\rangle$ for all e in $D(T)$. We define $T^*(f^*) = e^*$. The domain of T^* is weak*-dense in F^* and T^* is a closed linear transformation from F^* to E^*.

In stating the results of this section we shall use two topologies on E and F, the strong and the weak, and two topologies on E^* and F^*, the strong and the weak. A transformation T from $D(T)$ into F is said to be a homomorphism (strong or weak) if it maps open sets in $D(T)$ (in the topology induced on $D(T)$

by the strong or weak topologies on E, respectively) onto open sets in $R(T)$ (in the topology induced on $R(T)$ by the strong or weak topologies of F, respectively). Similarly, we define T^* to be a strong or weak* homomorphism from $D(T^*)$ into E^* by the requirement that T^* maps open sets of $D(T^*)$ in the appropriate topology onto open sets of $R(T^*)$ in the appropriate topology. An equivalent condition for T to be a strong homomorphism is that there exists a constant c such that for each f in $R(T)$, there exists e in $D(T)$ with $\|e\| \leq c\|f\|$, $Te = f$. (When no topology is explicitly mentioned, the strong topology is indicated.)

Theorem 1.1: Let E and F be two Banach spaces, T a closed linear operator with dense domain $D(T)$ in E and range $R(T)$ in F. Then the following properties of T and its adjoint operator T^* are equivalent to one another:

(a) $R(T)$ is closed in F,
(b) $R(T^*)$ is closed in E^*,
(c) $R(T^*)$ is weak*-closed in E^*,
(d) T is a homomorphism of $D(T)$ into F,
(e) T is a weak homomorphism of $D(T)$ into F,
(f) T^* is a homomorphism of $D(T^*)$ into E^*,
(g) T^* is a weak homomorphism of $D(T^*)$ into E^*,
(h) $R(T) = \{f: \langle f, f^*\rangle = 0$ for f^* in $N(T^*)\}$,
(i) $R(T^*) = \{e^*: \langle e, e^*\rangle = 0$ for all e in $N(T)\}$,
(j) $R(T)$ is of second category on itself,
(k) $R(T^*)$ is of second category on itself.

For bounded operators, Theorem 1.1 is due to Banach [1], Chap. X, pp. 145—150. A generalization of Banach's results to Fréchet spaces (i.e. complete locally convex metrizable linear spaces) was given by Dieudonné and Schwartz in [9]. We give a similar generalization of their results for closed operators in Fréchet spaces in the following:

Theorem 1.2: Let E and F be Fréchet spaces, T a closed linear operator with dense domain $D(T)$ in E and range $R(T)$ in F. Then the following properties of T and its adjoint operator T^* are all equivalent:

(a) $R(T)$ is closed in F,
(c) $R(T^*)$ is weak*-closed in E^*,
(d) T is a homomorphism of $D(T)$ into F,
(e) T is a weak homomorphism of $D(T)$ into F,
(f) T^* is a weak* homomorphism of $D(T^*)$ into E^*,
(h) $R(T) = \{f: \langle f, f^*\rangle = 0$ for all f^* in $N(T^*)\}$,
(i) $R(T^*) = \{e^*: \langle e, e^*\rangle = 0$ for all e in $N(T)\}$,
(j) $R(T)$ is of second category on itself.

The essential part of Theorem 1.1 is the equivalence of conditions (a) and (b)[1]. The proofs of both Theorems 1.1 and 1.2 follow from the more general result of Theorem 2.1 of Section 2. We shall give in the present section

[1]) The writer has been informed by J. T. Schwartz that the equivalence of (a) and (b) for closed operators in Banach spaces was established by G. C. Rota in his Yale doctoral dissertation in 1956.

a more elementary proof of this equivalence for reflexive Banach spaces, and discuss various corollaries and applications of the general result. We remark that in Hilbert space, the equivalence of (a) and (b) of Theorem 1 was remarked without explicit proof by Visik in [28], and an important special case of the result was applied by Visik in [29] to the study of mixed initial-boundary value problems.

Theorem 1.1′: Let E and F be Banach spaces, E and F reflexive, T a closed linear transformation with dense domain in E and range in F, T^* its adjoint operator. Then $R(T)$ is closed iff $R(T^*)$ is closed.

Proof: We begin by showing that if $R(T)$ is closed in F, then $R(T^*)$ must be closed in E^*. Let $F_1 = R(T)$. Then T may be considered as the composition jT_1 of a mapping T_1 of E onto F_1 with the injection mapping j of F_1 of F. It follows easily from the Hahn-Banach theorem that j^* maps F^* onto F_1^*, and that $R(T^*) = R(T_1^*)$. Thus it suffices to consider the case that $R(T) = F$. Then T^* is one-to-one, and by the Banach open mapping theorem, there exists a constant $c > 0$ such that for each f in F, there exists an e in $D(T)$ with $Te = f$, $\|e\| \leq c\|f\|$. Then for each f^* in $D(T^*)$, we have

$$|\langle f, f^*\rangle| = |\langle Te, f^*\rangle| = |\langle e, T^*f^*\rangle| \leq \|e\| \cdot \|T^*f^*\| \leq c\|f\| \cdot \|T^*f^*\| .$$

Letting f be an arbitrary element of F with $\|f\| = 1$, we obtain

$$\|f^*\| \leq c\|T^*f^*\| ,$$

i.e. $\|(T^*)^{-1}\| \leq c$. Since T^* is closed, $(T^*)^{-1}$ is closed, and, since it is bounded, its domain, which is $R(T^*)$, must be closed in E^*.

Suppose now that $R(T^*)$ is closed. Since E and F are reflexive, $E^{**} = E$, $F^{**} = F$, and $T^{**} = T$. It follows from the preceeding portion of the argument that $R(T) = R((T^*)^*)$ is closed.

Another result for closed operators which we shall find useful is the following generalization of the classical theorem of F. Riesz for transformations of the form $I + C$, C completely continuous, which was given for bounded transformations in Banach spaces by B. Yood [31] and extended to continuous linear mappings of Fréchet spaces by L. Schwartz [25].

Theorem 1.3: Let T be a closed linear transformation with domain $D(T)$ dense in the Fréchet space E and range in the Fréchet space F. Let C be a T-compact linear operator from E to F, i. e. $D(C) \supset D(T)$, while there exist open neighborhoods U and V of null in E and F such that C maps $U \cap T^{-1}(V)$ into a compact subset of F. (In the case of Banach spaces, this last condition becomes: C maps each set $\{e: \|e\| + \|Te\| \leq r\}$ into a compact set.)

Suppose that $R(T)$ is a closed subspace with finite defect in F. Then for $T + C$, which is a closed operator with domain $D(T)$, $R(T + C)$ is a closed subspace with finite defect in F.

If in addition, $N(T)$ is of finite dimension, then $N(T + C)$ is of finite dimension.

Theorem 1.3 will be established as a consequence of Theorem 2.3 of Section 2.

As a simple corollary of Theorem 1.2, we have the following theorem (which is established for Banach spaces by a direct argument in § 3 of Chap. II of HILLE and PHILLIPS [13]).

Theorem 1.4: Let T be a closed linear transformation with dense domain in the Fréchet space E and range in the Fréchet space F. Then:

(i) T has a continuous inverse iff $R(T^*) = E^*$.

(ii) If $R(T)$ is of the second category in F, then $R(T) = F$ and T is a homomorphism.

(iii) The equation $Te = f$ has one and only one solution e in $D(T)$ for each f in F iff the equation $T^* f^* = e^*$ has one and only one solution f^* in $D(T^*)$ for each e^* in E^*.

Proof: (i) If T has a continuous inverse, then $R(T)$ is closed and T is one-to-one. By Theorem 1.2, $R(T^*) = E^*$. If $R(T^*) = E^*$, on the other hand, $R(T^*)$ is weak* closed. By Theorem 1.2, $R(T)$ is closed, $N(T) = \{0\}$, and T is a homomorphism. T^{-1} is then continuous.

(ii) If $R(T)$ is of the second category in F, it must be of the second category on itself. By Theorem 1.2, $R(T)$ is closed in F and T is a homomorphism. If $R(T)$ were a proper closed subspace of F, $R(T)$ would be nowhere dense. Thus $R(T) = F$.

(iii) The equation $Te = f$ has one and only one solution for each f in F iff $R(T) = F$, $N(T) = \{0\}$. But then $R(T^*) = E^*$, and $N(T^*) = \{0\}$, by Theorem 1.2 and the equation $T^* f^* = e^*$ has one and only one solution for each e^* in E^*. The converse implication follows similarly.

Theorem 1.5: Let T be a closed linear mapping with dense domain $D(T)$ in the Banach space E and range $R(T)$ in the Banach space F.

Suppose that either:

(a) There exists a mapping J of $D(T)$ into F^* and a positive constant c such that for all e in $D(T)$,

$$|\langle Te, Je\rangle| \geq c \, \|e\|_E \cdot \|Je\|_{F^*} \tag{1.1}$$

or

(b) There exists a continuous linear mapping J of $D(T)$ into F^* and a positive constant c such that for all e in $D(T)$

$$|\langle Te, Je\rangle| \geq c \, \|e\|_E^2 \tag{1.2}$$

or

(c) There exists a continuous linear mapping J of $D(T)$ into F^* and a positive constant c such that for all e in $D(T)$,

$$Re\,\langle Te, Je\rangle \geq c \, \|e\|_E^2 \, . \tag{1.3}$$

Then $R(T)$ is closed, $N(T) = \{0\}$, T^{-1} is a bounded mapping of $R(T)$ into E, and $R(T^*) = E^*$.

Proof: (c) obviously implies (b). On the other hand since $\|Je\|_{F^*} \leq c_1 \|e\|$, (b) implies (a). It suffices therefore to consider (a).

Since $|\langle Te, Je\rangle| \leq \|Te\|_F \cdot \|Je\|_{F^*}$, (1.1) implies that $\|Te\|_F \geq c\,\|e\|_E$. Thus $N(T) = \{0\}$, T^{-1} is bounded on $R(T)$, and $R(T)$ is closed. It follows from Theorem 1.1 that $R(T^*) = E^*$.

Theorem 1.6: Let T be a closed linear transformation with dense domain in the Banach space E and range $R(T)$ in the Banach space F, T^* its adjoint operator. Suppose that one of the conditions (a), (b), or (c) of Theorem 1.5 holds and that $N(T^*) = \{0\}$. Then $R(T) = F$, and the equations $Te = f$, $T^* f^* = e^*$ have one and only one solution for each f in F and e^* in E^*.

Proof: By Theorem 1.5 and Theorem 1.1, $R(T) = \{f : \langle f, f^* \rangle = 0$ for all f^* in $N(T^*)\}$. Since $N(T^*) = \{0\}$, $R(T) = F$, and our conclusion follows.

Theorem 1.7: Let T be a closed operator from the Banach space E to the Banach space F, T^* its adjoint, J a continuous linear mapping of $D(T)$ into F^* such that $D(T^*) \subset R(J)$ and there exists a constant $c > 0$ for which

$$|\langle Te, Je \rangle| \geqq c \, \|e\|^2 .$$

Then the equation $Te = f$ has one and only one solution for each f in F (and a similar fact holds for the equation $T^* f^* = e^*$ for each e^* in E^*).

Proof: By Theorem 1.5 (b), $R(T)$ is closed, T is one-to-one, and $R(T^*) = F$. By Theorem 1.6, we need verify only that $N(T^*) = \{0\}$.

Suppose $f^* \in N(T^*)$. Since $D(T^*) \subset R(J)$, there exists e in $D(T)$ such that $Je = f^*$. For such an e, however,

$$0 = |\langle e, T^* f^* \rangle| = |\langle Te, Je \rangle| \geqq c \, \|e\|^2 .$$

We have $e = 0$, $f^* = Je = 0$, implying that $N(T^*) = \{0\}$.

Corollary to Theorem 1.7: Let H be a Hilbert space with inner product $[u, v]$, T a closed linear operator with dense domain $D(T)$ in H and range $R(T)$ in H, T^* its adjoint operator. Suppose that there exists a positive constant c such that

$$Re[Tu, u] \geqq c \, \|u\|^2$$

for all u in $D(T)$. Then $R(T^*) = H$.

If in addition, T is continuous, $R(T) = H$ and T^{-1} is a bounded operator on H.

Proof: We choose $Ju = u$. The first part of the corollary follows from Theorem 1.3, the second from Theorem 1.7.

For closed T, the result of Corollary 1.7 is due to VISIK [28], [29]. For bounded T, the stronger conclusion is also due to VISIK [27], but was first emphasized as a significant result in its own right by LAX and MILGRAM [17][2]). The use of more general operators J than the identity operator was begun in the work of LERAY [18] on the Cauchy problem for hyperbolic equations.

Theorem 1.8[3]): Let E and F be Banach spaces, V a locally convex linear space, T_1 and T_2 continuous linear mappings of V into E and F respectively, T_1^* and T_2^* the adjoint mappings of T_1 and T_2. Then: $R(T_1^*) \supset R(T_2^*)$ iff there exists a constant c_0 such that

$$\|T_2 v\| \leq c_0 \|T_1 v\| \tag{1.4}$$

for all v in V.

---

[2]) Other applications of this result have been made by the writer [5], [6], J. L. LIONS [18], [20], and L. NIRENBERG [21].

[3]) Theorem 1.8 is an extension of a result of FICHERA [10].

(ii) Suppose that $R(T_1)$ is dense in the Banach space E, and that (1.4) holds. Then for each pair of elements e^* in E^* and f^* in F^* for which $T_1^*(e^*) = T_2^*(f^*)$, i.e.

$$\langle T_1 v, e^*\rangle = \langle T_2 v, f^*\rangle, \quad v \in V, \tag{1.5}$$

we have

$$\|e^*\| \leq c_0 \|f^*\| . \tag{1.6}$$

Proof: Let E_1 be the closure of the range of T_1, F_1 the closure of the range of T_2. E_1 and F_1 are closed subspaces of E and F respectively, and we may consider T_1 and T_2 as mappings of V into E_1 and F_1, composed with the identity mappings j and k of E_1 and F_1 into E and F, respectively. By the Hahn-Banach theorem, j^* and k^* are open mappings of E^* onto E_1^* and of F^* onto F_1^*. Thus the ranges of T_1^* and T_2^* are not changed if we replace E and F by E_1 and F_1. Consequently, there is no loss of generality in assuming that $R(T_1)$ is dense in E, $R(T_2)$ is dense in F.

Under this latter assumption, T_1^* and T_2^* are injective mappings of E^* and F^* into V^*. Now if $R(T_1^*) \supset R(T_2^*)$, the mapping $Q'(f^*) = (T_1^*)^{-1}(T_2^* f^*)$ is well-defined for all f^* in E^* and by the continuity of T_2^* and T_1^*, Q' is a closed linear mapping of F^* into E^*. By the closed graph theorem Q' is bounded, i.e. $\|Q'(f^*)\| < c_0\|f^*\|$ for all f^* in F^*. For each v in V, however,

$$\langle T_2 v, f^*\rangle = \langle v, T_2^* f^*\rangle = \langle v, T_1^*(Q'f^*)\rangle = \langle T_1 v, Q'f^*\rangle .$$

Since $\|T_2 v\| = \sup\{|\langle T_2 v, f^*\rangle| : f^* \in F^*, \|f^*\| = 1\}$, the preceding equality implies

$$\|T_2 v\| \leq c_0 \|T_1 v\| . \tag{1.4}$$

On the other hand, suppose (1.4) holds. Then $N(T_1) \subset N(T_2)$, and the mapping Q defined by the assignment $Q(T_1 v) = T_2 v$ yields a well-defined and bounded linear transformation of $R(T_1)$ into F. Since $R(T_1)$ is dense in E, Q may be extended to a bounded linear operator from E into F. Let Q^* be the adjoint operator to Q. Q^* is obviously a continuous linear operator from F^* into E^*. Let $f^* \in F^*$, $v \in V$. We have

$$\langle v, T_1^*(Q^* f^*)\rangle = \langle T_1 v, Q^* f^*\rangle = \langle Q(T_1 v), f^*\rangle = \langle T_2 v, f^*\rangle = \langle v, T_2^* f^*\rangle .$$

Thus $T_1^*(Q^* f^*) = T_2^* f^*$, $R(T_2^*) \subset R(T_1^*)$, and (i) is proved.

For the proof of (ii), we observe that $T_1^*(e^*) = T_2^*(f^*)$ iff $e^* = Q^* f^*$, and the conclusion of (ii) follows from the boundedness of Q^*, which has already been established.

A consequence of Theorem 1.8 is the following result, which has been obtained by Fichera [11] in a slightly more special form for mappings into L^p-spaces. (Fichera's theorem makes the unnecessarily strong assumption that $R(L) = E$ in the notation of Theorem 1.9.)

Theorem 1.9: Let E and F be two Banach spaces, E^* and F^* their respective adjoint spaces, V a locally convex linear space. Let L, L', M, M' be continuous

linear mappings of V into E, F^*, F, and E^* respectively. (Cf. the following diagram)

$$\begin{tikzcd}[row sep=small] & E & & & F \\ V \arrow[ur, "L"] \arrow[dr, "M'"'] & & & V \arrow[ur, "M"] \arrow[dr, "L'"'] & \\ & E^* & & & F^* \end{tikzcd}$$

Suppose that for each pair of elements u and v of V, we have

$$(1.7) \qquad \langle Lu, M'v\rangle = \langle Mu, L'v\rangle .$$

Suppose further that $R(L)$ is dense in E, and that there exists a constant $c_0 > 0$ such that for all v in V

$$(1.8) \qquad \|Mv\| \leqq c_0\|Lv\| .$$

Then for each v of V, we have the inequality

$$(1.9) \qquad \|M'v\| \leqq c_0\|L'v\| .$$

Proof: Since $R(L)$ is dense in E, $(L^*)^{-1}$ exists on $R(L^*)$. By Theorem 1.8, $R(M^*) \subset R(L^*)$ and $Q^* = (L^*)^{-1}M^*$ is a bounded linear mapping of F^* into E^* with $\|Q^*\| \leqq c_0$.

Let u and v be elements of V, and set $f^* = L'v$. We have

$$\langle Mu, L'v\rangle = \langle u, M^*f^*\rangle = \langle u, L^*(Q^*f^*)\rangle = \langle Lu, Q^*f^*\rangle .$$

By (1.7), $\langle Lu, M'v\rangle = \langle Lu, Q^*f^*\rangle$. Since $R(L)$ is dense in E, $M'v = Q^*f^* = Q^*(L'v)$. By the bound on Q^*, we have $\|M'v\| \leqq c_0\|L'v\|$.

Section 2: Let E and F be locally convex linear Hausdorff spaces over the field of complex numbers. For the definition and general theory of locally convex linear spaces, we refer to BOURBAKI [2]. To be consistent with the usual Banach space conventions, our notation differs slightly from that of [2].

Our primary concern in this section is the theory of (not necessarily continuous) linear operators T with domain in E and range in F.

Let T be such an operator with domain $D(T)$, a linear subspace of E, and with range $R(T)$ in F. T is said to be a closed operator from E to F if the graph of T,

$$G(T) = \{(e, f) : e \in D(T), f = Te\}$$

is a closed subset of the product space $E \times F$. T is said to be closable as an operator from E to F if $cl(G(T))$ (the closure of $G(T)$ in $E \times F$) is the graph of a linear operator from E to F. The obvious necessary and sufficient condition for T to be closable is that $cl(G(T))$ include no elements of the form $(0, f)$ with $f \neq 0$.

Let E^* and F^* be the dual spaces of E and F respectively, which we shall assume, except when otherwise explicitly stated, to have the strong topology (i.e. uniform convergence on bounded subsets of E and F, respectively). For e in E, e^* in E^*, we shall write the pairing of e and e^* as $\langle e, e^*\rangle$. The linear space $F^* \times E^*$ is isomorphic with $(E \times F)^*$ under the pairing

$$\langle (e, f), (f^*, e^*)\rangle = \langle e, e^*\rangle + \langle f, f^*\rangle .$$

For a subset A of E, the polar of A is the subset A^0 of E^* consisting of those e^* for which $|\langle e, e^*\rangle| \leq 1$ for all e in A. If A is a subspace, the polar of A coincides with the annihilator of A, i.e. the set of e^* such that $\langle e, e^*\rangle = 0$ for all e in A. If B is a subset of E^*, the polar of B is the subset B^0 of E consisting of those e for which $|\langle e, e^*\rangle| \leq 1$ for all e^* in B. A subset A of the locally convex space E is said to be circled if αe lies in A for every e in A and each complex α with $|\alpha| \leq 1$. Then for each subset A of E, $A^{00} = (A^0)^0$ is the smallest closed convex circled set containing A. In particular, if A is a closed convex circled set, then $A^{00} = A$.

The weak*-topology on E^*, $\sigma(E^*, E)$ in the notation of [2], is the locally convex Hausdorff topology having as its base of neighborhoods at the origin the sets of the form $\{e_1, \ldots, e_k\}^0$, where $\{e_1, \ldots, e_k\}$ is any finite subset of E. The weak topology on E, $\sigma(E, E^*)$, is the locally convex Hausdorff topology having as its base of neighborhoods at zero, the sets of the form $\{e_1^*, \ldots, e_k^*\}^0$ where $\{e_1^*, \ldots, e_k^*\}$ is any finite subset of E^*. The Mackey topology on E, $\tau(E, E^*)$, is the locally convex Hausdorff topology having as its base of neighborhoods at zero, the sets of the form K^0, where K is any weak*-compact convex circled subset of E^*. The original (or strong) topology on E is finer than the weak topology and coarser than the Mackey topology. A convex subset A of E is closed in any one of these three topologies iff it is closed in each of the other two. If B is a subset of E^*, $B^{00} = (B^0)^0$ is the smallest weak*-closed convex circled subset containing B.

The locally convex linear space E is said to be a Mackey space if $\tau(E, E^*)$ is the strong topology on E. If E is metrizable, it is a Mackey space (Prop. 6, § 2, Chap. IV of [2]).

E is said to be fully complete [15] provided that a subset B of E^* is closed in the weak*-topology iff $B \cap K$ is weak*-compact for every weak*-compact subset K of E^* (the Krein-Smulian property for Banach spaces). In [15], J. L. Kelley has shown that full completeness of E is equivalent to the completeness of the uniform space of the closed linear subspaces of E with the Hausdorff uniformity. Every Fréchet space is fully complete [9].

Definition 2.1: E is said to satisfy condition (s) if for each closed subspace N of E and each compact convex circled subset K of E/N, there is a compact convex circled subset K_1 of E such that $\varphi(K_1) = K$, where φ is the canonical mapping of E onto E/N.

Every Fréchet space satisfies condition (s), [26].

Definition 2.2: E will be said to satisfy condition (t) if for every linear subspace M of E, the Mackey topology of M, $\tau(M, M^*)$, coincides with the topology induced on M by the Mackey topology of E, $\tau(E, E^*)$.

In general $\tau(M, M^*)$ is always finer than $\tau(E, E^*)$ restricted to M. A condition equivalent to condition (t) is that E^* with the weak* topology should satisfy condition (s).

If every linear subspace of E is a Mackey space, then E satisfies condition (t), since $\tau(M, M^*)$ coincides in that case with the induced strong topology on M,

which in turn is the topology induced by $\tau(E, E^*)$. Thus every metrizable space E satisfies condition (t).

A subset A of E is said to be absorbing if for each x in E there exists a complex number λ such that λx lies in A.

E is said to be tonnelé if every closed convex circled subset A of E which is absorbing has zero in its interior. Every Fréchet space and more generally every Baire space is tonnelé. If E is tonnelé, E is a Mackey space (Prop. 5, § 2, Chap. IV of [2]). Every quotient space of a tonnelé space E by a closed subspace N is also tonnelé. E is said to be fully tonnelé if every closed subspace of E is tonnelé. Every Fréchet space is fully tonnelé.

Pták [22] has shown that if E is fully complete, F tonnelé, every continuous linear mapping of E onto F is an open mapping. A. P. and W. Robertson [23] have proved a closed graph theorem which states that every mapping T of a tonnelé space F into a fully complete space E with a closed graph is continuous. Collins [8] has shown that if N is a closed subspace of a fully complete space E, then N and E/N are fully complete and also that fully complete spaces are complete.

If E_1 is a subspace of E, then restricting each element e^* of E^* to the subspace E_1, defines a continuous linear mapping φ of E^* onto E_1^*. In general φ is a weak* homomorphism (i.e. an open mapping in the weak* topology) of E^* on E_1^*, and defines a weak*-isomorphism of E^*/E_1^0 with E_1^*. However, φ need not be a homomorphism in the strong topology of E^* and E_1^*.

Definition 2.3: We shall say that a pair (E, E_1) satisfies condition (r) if Φ is a homomorphism of E^* on E_1^* (and, therefore, an isomorphism of E^*/E_1^0). If E^* is fully complete, E_1^* tonnelé, the pair (E, E_1) will satisfy condition (r).

Let T be a linear operator with its domain $D(T)$ in E and its range $R(T)$ contained in F. We shall suppose that $D(T)$ is dense in E. Let

$$D(T^*) = \{f^*: f^* \in F^*, \langle Te, f^*\rangle \text{ is continuous in } e \text{ on } D(T)\}.$$

Since $D(T)$ is dense, the canonical mapping φ of E^* on $\{D(T)\}^*$ is weak*-isomorphism onto. Therefore, for each f^* in $D(T^*)$, there exists an unique element e^* in E^* such that

$$\langle e, e^*\rangle = \langle Te, f^*\rangle$$

for all e in $D(T)$. We define $T^*f^* = e^*$. The operator T^* thus defined with domain $D(T^*)$ contained in F^* and range $R(T^*)$ in E^* is called the adjoint operator of T.

Let $G(-T)$ be the graph of the operator $(-T)$, $G(T^*)$ the graph of T^*. We note that by the definition of T^*, the element (f^*, e^*) of $F^* \times E^*$ will lie in $G(T^*)$ iff $\langle Te, f^*\rangle = \langle e, e^*\rangle$ for all e in $D(T)$. This last equation may be re-written $\langle (e, f), (f^*, e^*)\rangle = 0$ for all (e, f) in $G(-T)$.

Lemma 2.1: (i) T^* is a closed operator from F^* to E^*. Indeed $G(T^*)$ is a weak*-closed subset of $F^* \times E^*$.

(ii) For e in E, f in F, the element (e, f) of $E \times F$ lies in $cl(G(T))$ iff $\langle e, T^*f^*\rangle = \langle f, f^*\rangle$ for all f^* in $D(T^*)$. In particular when T is closed, e lies in $D(T)$ with $Te = f$ iff $\langle e, T^*f^*\rangle = \langle f, f^*\rangle$ for all f^* in $D(T^*)$.

(iii) T is closable iff $D(T^*)$ is weak*-dense in E^*.

Proof: (i) Since $G(T^*) = G(-T)^0$, $G(T^*)$ is weak*-closed and therefore strongly closed in $F^* \times E^*$.

(ii) $cl(G(T))$ is a closed convex circled subset of $E \times F$. Since $G(-T^*) = G(T)^0$, we have $c\,l(G(T)) = c\,l(G(T))^{00} = G(T)^{00} = G(-T^*)^0$. In other word, (e, f) lies in $c\,l(G(T))$ iff $\langle e, T^* f^*\rangle - \langle f, f^*\rangle = 0$ for all f^* in $D(T^*)$, i.e. $\langle e, T^* f^*\rangle = \langle f, f^*\rangle$.

(iii) T is closable iff there exists no element of the form $(0, f)$ in $c\,l(G(T))$ with $f \neq 0$. Suppose such an element $(0, f)$ exists. Then for each f^* in $D(T^*)$, since the relation $\langle (e, f), (f^*, -T^* f^*)\rangle = 0$ for (e, f) in $G(T)$ persists by continuity for (e, f) in $c\,l(G(T))$, we have $0 = \langle 0, -T^* f^*\rangle + \langle f, f^*\rangle = \langle f, f^*\rangle$. Thus $D(T^*)$ is contained in $\{f\}^0$, and $D(T^*)$ cannot be weak*-dense. On the other hand, if $D(T^*)$ is not weak*-dense, since F is the dual space of F^* taken with the weak*-topology, there exists an f in F such that $\langle f, f^*\rangle = 0$ for all f^* in $D(T^*)$, $f \neq 0$. But then $\langle 0, T^* f^*\rangle = \langle f, f^*\rangle$ for all f^* in $D(T^*)$, which by (ii) implies $(0, f) \in c\,l(G(T))$ and T is not closable.

Let t_1 be a topology on E, t_2 a topology on F. T is said to be a (t_1, t_2) homomorphism if $T(A)$ is an open set in $R(T)$ in the topology induced by t_2 for every subset A of $D(T)$ open in the topology induced by t_1. T is said to be a (t_1, t_2) isomorphism if T is an injective, surjective (t_1, t_2) homomorphism of $D(T)$ on $R(T)$. As usual, if no topologies are specified, a homomorphism will refer to a homomorphism in the strong topologies on E and F. If t_1 and t_2 are the weak topologies we shall speak of a weak homomorphism. If t_1 and t_2 are the weak* topologies we shall speak of a weak* homomorphism.

Theorem 2.1: Let E and F be locally convex linear spaces, T a linear operator with dense domain $D(T)$ in E and range $R(T)$ in F. Let $G(T)$ be the graph of T considered as a subspace of $E \times F$, T^* the adjoint operator to T with domain $D(T^*)$ in F^* and range $R(T^*)$ in E^*, $N(T)$ and $N(T^*)$ the null spaces of T and T^* respectively.

Consider the following properties of T and T^*:

(a) $R(T)$ is closed in F,
(b) $R(T^*)$ is closed in E^*,
(c) $R(T^*)$ is weak*-closed in E^*,
(d) T is a homomorphism of $D(T)$ into F,
(e) T is a weak homomorphism of $D(T)$ into F,
(f) T^* is a homomorphism of $D(T^*)$ into E^*,
(g) T^* is a weak* homomorphism of $D(T^*)$ into E^*,
(h) $R(T) = \{f : \langle f, f^*\rangle = 0$ for all f^* in $N(T^*)\}$,
(i) $R(T^*) = \{e^* : \langle e, e^*\rangle = 0$ for all e in $N(T)\}$,
(j) $R(T)$ is tonnelé.

Then the following relations hold between these properties:

(I) $(a) \Leftrightarrow (h) \Leftrightarrow (g)$
$(b) \Leftarrow (c) \Leftrightarrow (i) \Leftrightarrow (e) \Leftarrow (d)$.

(II) If $G(T)$ is fully complete and F is fully tonnelé,

$$\begin{array}{c} (j) \Longleftrightarrow (a) \Longleftrightarrow (h) \Longleftrightarrow (g) \Longleftrightarrow (d) \\ \Downarrow \\ (b) \Leftarrow (c) \Longleftrightarrow (i) \Longleftrightarrow (e) \end{array}$$

(III) If $G(T)$ is fully complete, F fully tonnelé and satisfying condition (t), then

$$\begin{array}{c} (j) \Longleftrightarrow (a) \Longleftrightarrow (h) \Longleftrightarrow (g) \Longleftrightarrow (d) \\ \Updownarrow \\ (b) \Leftarrow (c) \Longleftrightarrow (i) \Longleftrightarrow (e). \end{array}$$

(IV) If $G(T)$ is fully complete, F fully tonnelé, and the pair $(E \times F,\ G(T))$ satisfies condition (r), then $(f) \Rightarrow (c)$.

(V) If $G(T)$ and F^* are fully complete, while F and $G(T^*)$ are fully tonnelé, and the pair $(E \times F,\ G(T))$ satisfies condition (r), then properties (a) to (j) are all mutually equivalent.

Proof: The assertion of Theorem 2.1 follows from Theorem 2.2 and Lemmas 2.2 to 2.9 which are proved below.

Theorem 2.2: Let T be a linear operator with domain $D(T)$, a subspace of E, and range $R(T)$ in F. Let G be the graph of T considered as a linear subspace of $E \times F$, S the continuous mapping of G into F defined by

$$S(e, Te) = Te .$$

Then:

(i) Properties (a), (c), (d), (e), (g), (h), (i), and (j) of Theorem 2.1 for T as a mapping of E into F are equivalent to the corresponding properties for S as a mapping of G into F.

(ii) Properties (b) and (f) for S as a mapping of G into F imply properties (b), (f) and (h) for T as a mapping from E into F.

(iii) If the pair $(E \times F,\ G(T))$ has property (r), then properties (b) and (f) for T are equivalent to the corresponding properties for S.

Proof: Let J be the mapping of $D(T)$ into G defined by $Je = (e, Te)$, J_1 the mapping of G into E defined by $J_1(e, Te) = e$. J is a homomorphism, J_1 is continuous, and $J_1 Je = e$ for e in $D(T)$.

We shall designate properties $(a), \ldots, (j)$ for S by $(a)_S, \ldots, (j)_T$, and the properties $(a), \ldots, (j)$ for T by $(a)_T, \ldots, (j)_T$. Since (h) is equivalent to (a) and (i) to (c), we need not consider them explicitly. We remark first that $R(T) = S(G)$ so that $(a)_S$ is equivalent to $(a)_T$ and $(j)_S$ to $(j)_T$. Furthermore, $T = SJ$, where J is a homomorphism and hence a weak homomorphism. Thus $(d)_S$ is equivalent to $(d)_T$, and $(e)_S$ is equivalent to $(e)_T$.

Let φ be the canonical mapping of $F^* \times E^*$ on G^*. By a preceeding remark, φ is continuous and a weak*-homomorphism, which induces a weak*-isomorphism of $(F^* \times E^*)/G(-T^*)$ onto G^*. [Note that $G^0 = G(T)^0 = G(-T^*)$.]

Since S is a continuous mapping of G into F, S^* is a continuous mapping of F^* into G^*. Let f^* be an element of F^*. Then for each (e, Te) in G, we have

$$\langle (e, Te), S^* f^* \rangle = \langle S(e, Te), f^* \rangle = \langle Te, f^* \rangle = \langle (e, Te), (f^*, 0) \rangle .$$

Thus $S^*f^* = \varphi(f^*, 0)$. Let $R(S^*)$ be the range of S^*. Since φ is a weak* homomorphism of $F^* \times E^*$ onto G^*, $R(S^*)$ is weak*-closed in G^* iff $\varphi^{-1}(R(S^*))$ is weak*-closed in $F^* \times E^*$. But the elements of $\Phi^{-1}(R(S^*))$ are precisely those of the form $(f^*, 0) + (f_1^*, -T^*f_1^*)$, where f^* is any element of F^* and f_1^* any element of $D(T^*)$. Since for fixed f_1^*, $f^* + f_1^*$ can be made equal to an arbitrary element of F^* by suitable choice of f^*, $\varphi^{-1}(R(S^*)) = F^* \times R(-T^*)$, which is weak*-closed in $F^* \times E^*$ iff $R(T^*)$ is weak*-closed in E^*. Thus $(c)_S$ is equivalent to $(c)_T$.

Since J_1 is a continuous mapping of G into E, J_1^* is a continuous mapping of E^* into G^*. Let $e^* \in E^*$, $(e, Te) \in G$. Then

$$\langle (e, Te), J_1^* e^* \rangle = \langle J_1(e, Te), e^* \rangle = \langle e, e^* \rangle = \langle (e, Te), (0, e^*) \rangle .$$

Thus $J_1^* e^* = \varphi(0, e^*)$. Now, we note that $J_1^* e^* = S^* f^*$ for e^* in E^*, f^* in F^* iff

$$(0, e^*) - (f^*, 0) = (-f_1^*, T^* f_1^*)$$

for some f_1^* in $D(T^*)$, i.e. $(f_1^*, -e^*) = (f^*, -T^*f_1^*)$. This last equation is equivalent to saying that f^* lies in $D(T^*)$ and $T^*f^* = e^*$. It follows that for every set A in F^*, $T^*(A \cap D(T^*)) = (J_1^*)^{-1}(S^*(A))$. J_1^* is continuous and hence weak*-continuous. Thus if $S^*(F^*)$ is closed in G^*, $R(T^*)$ is closed in E^*, so that $(b)_S$ implies $(b)_T$. If A is an open set in F^*, $T^*(A \cap D(T^*)) = (J_1^*)^{-1}(S^*(A))$ is open in the range of T^* if $S^*(A)$ is open in $S^*(F^*)$, so that $(f)_S$ implies $(f)_T$. Similarly $(g)_S$ implies $(g)_T$.

Suppose that $(g)_T$ holds. To establish $(g)_S$, we must show that if A is a weak*-open subset of F^*, then $S^*(A)$ is a weak*-open subset of $R(S^*)$. Since φ is a weak*-homomorphism of $F^* \times E^*$ on G^*, it suffices to show that $\varphi^{-1}(S^*(A))$ is a weak*-open subset of $\varphi^{-1}(R(S^*))$. We have already verified that $\varphi^{-1}(R(S^*)) = F^* \times R(-T^*)$. By a similar calculation, we observe that $\varphi^{-1}(S^*(A))$ consists of all elements of $F^* \times E^*$ of the form $(a^* + f_1^*, -T^*f_1^*)$ with a^* in A, f_1^* in $D(T^*)$. Let $(a^* + f_1^*, -T^*f_1^*)$ be a fixed such element, and choose a weak* neighborhood V of zero in F^* such that $a^* + v_1^* - v_2^* \in A$ for any v_1^*, v_2^* in V. Since T^* is a weak* homomorphism of $D(T^*)$ into E^*, by assumption, there exists a convex circled weak* neighborhood U of zero in E^* such that for u^* in $U \cap R(T^*)$, there exists v^* in $V \cap D(T^*)$ such that $T^*v^* = -u^*$. If a^* is any element of A, $(a^* + f_1^* + v^*, -Tf_1^* + u^*)$ lies in $\varphi^{-1}(S^*(A))$. Let v_1^* be an arbitrary element of V. Then $a^* = a_0^* + v_1^* - v^*$ lies in A, and $(a_0^* + f_1^* + v_1^*, -T^*f_1^* + u^*)$ lies in $\varphi^{-1}(S^*(A))$. But $(a_0^* + f_1^* + v_1^*, -T^*f_1^* + u^*)$ is an arbitrary element of $(a_0^* + f_1^* + V) \times ((-T^*f_1^* + U) \cap R(-T^*))$ which is a weak* neighborhood of $(a_0^* + f_1^*, -T^*f_1^*)$ in $F^* \times R(-T^*)$. Thus $\varphi^{-1}(S^*(A))$ is weak* open in $R(S^*)$ and $(g)_S$ follows.

The proofs given above for $(g)_S$ as a consequence of $(g)_T$ and $(c)_S$ as a consequence of $(c)_T$ would yield the equivalence of $(b)_S$ and $(b)_T$ as well as $(f)_S$ and $(f)_T$ provided that φ were a homomorphism of $F^* \times E^*$ onto G^* in the strong topology. The hypothesis of (iii) is precisely that φ is a homomorphism in the strong topology, and the conclusion of (iii) follows under that assumption.

Corollary to Theorem 2.2: Suppose that $F^* \times E^*$ is fully complete and G^* is tonnelé. Then properties (a)—(j) for S are equivalent to properties (a)—(j) for T.

Proof: Since each continuous linear transformation of the fully complete space $F^* \times E^*$ onto a tonnelé space G^* is a homomorphism, the pair $(E \times F, G)$ satisfies condition (r), and our conclusion follows from Theorem 2.2.

Lemma 2.2: Let E and F be locally convex linear spaces, T a continuous linear mapping of E into F, T^* the dual mapping of F^* into E^*. Then properties (a) and (h) are equivalent to property (g) for T, while properties (c) and (i) are equivalent to property (e) for T.

Proof: Since weak closure and strong closure are the same for subspaces of F, (a) is equivalent to the property that $R(T)$ is a weakly closed subset of F. The equivalence of (c) and (e) follows from Prop. 4, § 4, Chap. IV of [2], since every strongly continuous linear mapping is also weakly continuous. The equivalence of (a) and (g) follow from the same proposition combined with the fact that the dual of E^* with the weak* topology is E, the dual of F^* with the weak* topology is F, and the dual of T^* as a weak* continuous mapping of F^* into E^* is our original mapping T.

Lemma 2.3: Let E be a fully complete linear space, F fully tonnelé, and T a continuous linear mapping of E into F. Then (a) is equivalent to (d) for T.

Proof: Since T is a continuous linear mapping of the fully complete space E onto the closed linear subspace $R(T)$ of F, which by the hypothesis is tonnelé, it follows from the result of Pták stated above that T is a homomorphism. Thus (a) implies (d).

On the other hand, if E is fully complete, N the null space of the homomorphism T, then N is a closed subspace of E and E/N is fully complete by [8]. Since every fully complete space is complete, E/N is complete and T induces an isomorphism of E/N onto $R(T)$. Thus $R(T)$ is complete, and therefore must be a closed subspace of F.

Lemma 2.4: Let T be a continuous mapping of the locally convex linear space E into the locally convex linear space F. Then (d) implies (e) for T. In addition (c) implies (b) for T.

Proof: The fact that (d) implies (e) for T is a direct consequence of Prop. 9, § 4, Chap. IV of [2]. Since every weak*-closed set in E^* is closed, (c) implies (b).

Lemma 2.5: Let E be fully complete, F fully tonnelé, T a continuous linear mapping of E into F. Then:

(i) If T^* is an isomorphism of F^* into E^*, then $R(T^*)$ is weak* closed in E^*, i.e. $(c)_T$ holds.

(ii) $(f)_T$ implies $(c)_T$.

Proof: (i) Since E is fully complete, $R(T^*)$ is weak* closed in E^* iff $R(T^*) \cap K$ is weak*-compact for each weak*-compact subset K of E^*. Every weak*-compact subset of E^* is bounded (Prop. 7, § 3, Chap. III of [2]). Since T^* is an isomorphism of F^* into E^* which is also weak*-continuous, $(T^*)^{-1}(K)$ is a bounded weak*-closed subset of F^*. Since F is tonnelé, $(T^*)^{-1}(K)$ is weak*-

compact in F^* (Theorem 3, § 2, Chap. IV of [2]). Hence, $R(T^*) \cap K = T^*((T^*)^{-1}(K))$ is weak*-compact in E^*. The weak*-closedness of $R(T^*)$ is thus established.

(ii) Let $M = cl(R(T))$. The canonical mapping φ of F^* into M^* is continuous and a weak*-homomorphism. Let T_1 be the mapping of E into M induced by T, T_1^* the adjoint mapping of M^* into E^*. Then, $T^* = T_1^* \varphi$. Suppose that T^* is a homomorphism. Then for every open subset A of M^*, $T_1^*(A) = T^*(\varphi^{-1}(A))$ is open in $R(T_1^*) = R(T^*)$, and T_1^* is a homomorphism. But the range of T_1 is a dense subset of M, so that T_1^* is injective and an isomorphism. Applying the conclusion of (i), $R(T_1^*) = R(T^*)$ is weak*-closed in E^*.

Lemma 2.6: Let E be fully complete, F a fully tonnelé space satisfying condition (t), T a continuous linear mapping of E into F. Then $(e)_T$ implies $(g)_T$.

Proof: Let T_1 be the mapping of E/N onto $R(T)$ induced by T, where N, the nullspace of T, is a closed linear subspace of E. It is sufficient to show that the mapping T_1, which is continuous and a weak isomorphism, is also an isomorphism. Let T_1^{-1} be the inverse mapping to T_1. T_1^{-1} is weakly continuous as a linear mapping from the weak topology of $R(T)$ (which is the topology induced on $R(T)$ by the weak topology of F) to the weak topology of E/N. Hence by Prop. 7, § 4, Chap. IV of [2], T_1^{-1} is continuous as a mapping from the Mackey topology $\tau(M, M^*)$ on M into the Mackey topology $\tau(E/N, (E/N)^*)$ on E/N, while the last topology is finer than the strong topology on E/N. Since F has property (t) and is tonnelé, the Mackey topology $\tau(M, M^*)$ is identical with the strong topology of M. Thus T_1^{-1} is continuous in the strong topologies, and $(g)_T$ is established.

Lemma 2.7: Let E^* be fully tonnelé, F^* fully complete, T a continuous linear mapping of E into F. Then $(b)_T$ implies $(f)_T$.

Proof: Suppose $R(T^*)$ is closed in E^*. T^* is a continuous linear mapping of the fully complete space F^* onto $R(T^*)$, and $R(T^*)$ is tonnelé, since it is a closed subspace of the fully tonnelé space E^*. Thus by the Pták open mapping theorem, T^* is a homomorphism of E^* on $R(T^*)$ and into E^*, and $(f)_T$ is established.

Lemma 2.8: Let E be fully complete, F fully tonnelé, E^* fully tonnelé, F^* fully complete, T a continuous linear mapping of E into F. Then $(b)_T$ implies $(a)_T$.

Proof: Let $M = cl(R(T))$. T may be considered as inducing a continuous linear mapping T_1 of E into M, and the range of T_1^* is equal to $R(T^*)$. By Lemma 2.7, if $(b)_T$ holds, then T^* is a homomorphism of F^* into E^*. If φ is the canonical mapping of F^* onto M^*, φ is continuous and $T^* = T_1^* \varphi$. It follows that T_1^* is also a homomorphism. Since $R(T_1)$ is dense in M, T_1^* is an injective mapping. We may assume without loss of generality in the remaining part of the proof that $M = F$.

Let f be an arbitrary element of F. We wish to show that f lies in $R(T_1)$. Let $H = \{f\}^0$, K an arbitrary weak*-compact subset of E^*. H is a weak*-

closed subset of F^*. $T_1^*(H) \cap K$ is compact in the weak* topology and hence strongly bounded. Since T_1^* is an isomorphism of M^* into E^*, $(T_1^*)^{-1}(T_1^*(H) \cap K) = H \cap (T_1^*)^{-1}(K)$ is bounded and weak*-closed in F^*. Since F is tonnelé, $H \cap (T_1^*)^{-1}(K)$ is weak*-compact. Since T_1^* is weak*-continuous, $T_1^*(H) \cap K$ is weak*-compact. Since this is true for an arbitrary weak*-compact K and since E is fully complete, $T_1^*(H)$ is weak*-closed.

By the Hahn-Banach theorem, there exists an element f_0^* in F^* which is not contained in H. Since T_1^* is injective, $T_1^*(f_0^*)$ does not lie in $T_1^*(H)$. Since the dual space of E^* with the weak* topology is E with a suitable topology, by the Hahn-Banach theorem there exists an e in E such that $\langle e, T_1^* h\rangle = 0$ for h in H, while $\langle e, T_1^*(f_0^*)\rangle \neq 0$. Every element f^* of F^* may be written in the form $a(f^*) f_0^* + h$, where h lies in H and $a(f^*) = \langle f, f^*\rangle / \langle f, f_0^*\rangle$. Thus,

$$\langle T_1 e, f^*\rangle = \langle e, T_1^*(h)\rangle + a(f^*)\langle e, T_1^*(f_0^*)\rangle = a_1^* \langle f, f^*\rangle ,$$

where $a_1^* = \langle e, T_1^*(f_0^*)\rangle / \langle f, f_0^*\rangle \neq 0$. It follows that $T_1^{-1}(a_1^{-1} e) = f$, and $(a)_T$ is established.

Lemma 2.9: Let E be fully complete, F fully tonnelé, T a continuous linear mapping of E into F. Then $(j)_T$ is equivalent to $(a)_T$.

Proof: If $R(T)$ is a closed subspace of the fully tonnelé space F, it must be tonnelé. Conversely, if $R(T)$ is tonnelé, the Ptak open mapping theorem asserts that T is a homomorphism of the fully complete space E on $R(T)$. Thus $R(T)$ is fully complete, hence complete, hence closed in F.

Theorem 2.3: Let E be a fully complete space satisfying condition (s), F a fully tonnelé space, T_0 a continuous linear mapping of E into F, C a compact mapping of E into F. Suppose that $R(T_0)$ is a closed subspace of F with finite co-dimension in F.

Then if $T = T_0 + C$, $R(T)$ is a closed subspace of F with finite co-dimension in F. If in addition $N(T_0)$ is of finite dimension, then so is $N(T)$.

Proof: Since F is fully tonnelé and $R(T_0)$ is closed, we know that $R(T_0)$ is tonnelé. Since E is fully complete, T_0 is a homomorphism of E into F, and hence a weak homomorphism. In particular, $R(T_0^*)$ is a weak*-closed subset of E^*. Furthermore, $N(T_0^*) = R(T_0)^0$, and by the hypothesis on the co-dimension of $R(T_0)$, $N(T_0^*)$ has finite dimension.

We shall denote by ξ the topologies on E^* and F^*, respectively, with bases of closed neighborhoods of the origin K^0 : K convex and compact. By Theorem 1 of [16], C^* is a compact mapping from the ξ-topology on F^* into the ξ-topology on E^*.

Let φ be the canonical mapping of E onto $E/N(T_0)$, ψ the injection mapping of $R(T_0)$ into F. T_0 may be written as the composition $\psi S_1 \varphi$, where S_1 is the isomorphism of $E/N(T_0)$ onto $R(T_0)$ induced by T_0. Thus, $T_0^* = \varphi^* S^* \psi^*$. Since S_1 is an isomorphism, S_1^* is an isomorphism in the ξ-topologies on $R(T_0)^*$ and $(E/N(T_0))^*$. Since φ^* and ψ^* are continuous in the strong topology, they are continuous in the ξ-topologies.

Let P be a continuous idempotent linear mapping of F onto $R(T_0)$. Such a mapping exists by the Hahn-Banach theorem and the finite

co-dimension of $R(T_0)$ in F.) If K is a compact subset of F, let $K_1 = P(K)$. If $\beta \in K_1^0$, let $\lambda = P^* \beta$. Then, $|\langle f, \lambda\rangle| = |\langle f, P^*\beta\rangle| = |\langle Pf, \beta\rangle| \leq 1$ for f in K, i.e. $\lambda \in K^0$. Moreover, $\psi^*(\lambda) = \beta$. Thus $\psi^*(K^0) \supset K_1^0$, and ψ^* is a ξ-homomorphism of F^* onto $R(T_0)^*$.

Similarly, since E satisfies condition (s), for every compact convex subset K_1 of $E/N(T_0)$ there exists a compact set K in E such that $\varphi(K) = K_1$. Since E is fully complete and hence complete, K may be chosen convex. Then $\beta \in K^0 \cap R(\varphi^*)$ iff $|\langle e, \beta\rangle| \leq 1$ for e in K and $\langle e, \beta = \langle \varphi e, \lambda\rangle$ with $\lambda \in (E/N(T_0))^*$. But $|\langle e, \beta\rangle| = |\langle \varphi e, \lambda\rangle| \leq 1$ for e in K is equivalent to the assertion that $\lambda \in K_1^0$, i.e. $\beta \in \varphi^*(K_1^0)$. Thus φ^* is a ξ-homomorphism of $(E/N(T_0))^*$ into E^*.

Finally, therefore, T_0^* is a ξ-homomorphism of F^* into E^* with finite dimensional nullspace, C^* a ξ-compact linear mapping of F^* into E^*, and the range of T_0^*, being weak*-closed in E^*, is also ξ-closed in E^*. By Theorem 8 of [16], $T^* = T_0^* + C^*$ is a ξ-homomorphism of F^* onto a ξ-closed subspace of E^* with finite co-dimension in E^*.

By Theorem 2, § 2, Chap. IV of [2], the duals of E^* and F^* with the ξ-topology are E and F, respectively, with some suitable topology. It follows that every ξ-closed subset of E^* is also weak*-closed and that every ξ-homomorphism of F^* into E^* is a weak* homomorphism. Thus T^* is a weak* homomorphism with a weak*-closed range. But by Prop. 4, § 4, Chap. IV of [2], T is then a weak homomorphism and $R(T)$ is a closed subspace of F with finite co-dimension in F.

Section 3: In the present section we shall give a more general treatment of the abstract theory of boundary problems previously discussed by Visik [28], Hörmander [14], and Visik and Ladizenskaja [30]. The proofs which we give below for our results in a more general framework are simpler when specialized to the original formulation than those given in the previous dsicussions.

Let E and F be two locally convex linear spaces, T_0 a linear transformation with dense domain $D(T_0)$ in E and range $R(T_0)$ in F. If T_0' is a linear transformation with domain $D(T_0')$ in F^* and range $R(T_0')$ in E^*, T_0' is said to be a formal adjoint of T_0 iff for all e in $D(T_0)$ and f^* in $D(T_0')$,

$$\langle T_0 e, f^*\rangle = \langle e, T_0' f^*\rangle. \tag{3.1}$$

There will exist at least one formal adjoint T_0' for T_0 with weak*-dense domain in F^* (namely, T_0^*) iff T_0 is closable.

We shall assume throughout that T_0 and T_0' are given with T_0' having a weak*-dense domain in F^*. Since T_0 and T_0' are closable, we shall assume them without loss of generality in our discussion to be closed linear transformations in their respective pairs of spaces.

We define T_1 as the restricted adjoint of T_0', i.e. $D(T_1) = \{e : e \in E;$ there exists f in F such that $\langle f, f^*\rangle = \langle e, T_0' f^*\rangle$ for all f^* in $D(T_0')\}$ with $T_1 e = f$ for e in $D(T_1)$. For each e in $D(T_1)$, $T_1 e$ is uniquely defined by the weak* denseness of $D(T_0')$. In addition, it follows from (3.1) that $T_0 \subseteq T_1$.

Definition 3.1: The closed densely defined linear operator T from E to F is said to be a realization operator for the pair (T_0, T_0') if $T_0 \subsetneq T \subsetneq T_1$.

The intuitive background for the definition of a realization operator is provided by the basic application of the definition: T_0 is defined by a linear partial differential operator L with domain $C_c^\infty(G)$, the family of infinitely differentiable functions with compact support in an open set G of Euclidean n-space, T_0' is defined by the formal adjoint differential operator L' restricted to $C_c^\infty(G)$, E and F are locally convex linear spaces such that E, F, E^*, and F^* can be continuously realized as linear subsets of the space of distributions on G with $C_c^\infty(G)$ dense in E and weak*-dense in F^* and with the pairing $\langle e, e^*\rangle$ and $\langle f, f^*\rangle$ for e in $C_c^\infty(G)$ and f^* in $C_c^\infty(G)$ being the distribution pairing. T_1 is then the differential operator L considered as defined for all e in E such that Le lies in F. Each realization operator T in this case is a closed restriction of the maximal realization T_1 such that the domain of T contains all of $C_c^\infty(G)$, the family of functions satisfying any null condition imposed on a function at the boundary of G. The realizations in this case correspond, therefore, to boundary-value conditions imposed on functions for which the differential operator L is defined in the sense of distributions.

Definition 3.2: The realization operator T is said to be a solvable operator for the pair (T_0, T_0') if T is a one-to-one mapping of $D(T)$ onto F, and if T^{-1} is a continuous linear mapping of F into E.

Definition 3.3: The realization operator T is said to be completely solvable for the pair (T_0, T_0') if T is solvable and if, in addition, T^{-1} is a compact linear mapping of F into E (i.e. maps some neighborhood of zero in F into a compact subset of E).

We shall treat two problems in the first part of this section: (I) To give a necessary and sufficient condition on the pair (T_0, T_0') for the existence of solvable and completely solvable operators T; (2) To classify solvable operators, if any exist.

We shall give three answers to Problem I, in order of increasing precision for Fréchet spaces, reflexive Banach spaces, and Hilbert spaces.

Theorem 3.1: Let E and F be two Fréchet spaces, T_0 a densely defined linear transformation in E with range in F, T_0' a linear transformation with weak* dense domain in F^* and range in E^* which is formally adjoint to T_0, T_1 the restricted adjoint of T_0'. Then a necessary and sufficient condition that there should exist a solvable operator for the pair (T_0, T_0') is that all the following three conditions should be satisfied:

(a) T_0 is one-to-one, and T_0^{-1} is a continuous mapping of $R(T_0)$ into E.

(b) T_0' is one-to-one, and $(T_0')^{-1}$ is a weak* continuous mapping of $R(T_0')$ into F^*.

(c) If $N(T_1)$ is the null space of T_1, $N(T_1) \times \{0\}$ has a closed complement M in $G(T_1)$ such that $G(T_0) \subset M$.

Theorem 3.2: Let E and F be two reflexive Banach spaces, T_0, T_0', and T_1 as in Theorem 3.1. Then a necessary and sufficient condition that there should exist a solvable operator for the pair (T_0, T_0') is that all of the following three conditions should be satisfied:

(a) T_0 is one-to-one and T_0^{-1} is a continuous mapping of $R(T_0)$ into E.
(b)' T_0' is one-to-one, and $(T_0')^{-1}$ is a continuous mapping of $R(T_0')$ into F^*.
(c) $N(T_1)\times\{0\}$ has a closed complement M in $G(T_1)$ such that $G(T_0)\subset M$.

Proof of Theorem 3.1: Let T_1 be the restricted adjoint of T_0'. The graph of T_1 is the set of elements (e, f) of $E\times F$ such that $\langle e, T_0' f^*\rangle = \langle f, f^*\rangle$ for all f^* in $D(T_0')$. If, we realize the graph of T_0' in $F^*\times E^*$, with the latter considered as the dual space of $E\times F$ under the pairing

$$\langle (e, f), (f^*, e^*)\rangle = \langle e, e^*\rangle - \langle f, f^*\rangle ,$$

then $G(T_1) = G(T_0')^0$. It follows that $G(T_1)^0$, which is in fact the graph of T_1^*, is the weak* closure of $G(T_0')$, i.e. T_1^* is the weak* closure of T_0'.

The necessity of conditions (a), (b), and (c) is obvious. Indeed, $T_0\subsetneq T$ implies that T_0^{-1} exists and is continuous on $R(T_0)$ satisfying (a). Furthermore $R(T) = F$ implies that $R(T_1) = F$, implying by Theorem 1.2 of Section 1 that T_0' is one-to-one and a weak* homomorphism. Thus $(T_0')^{-1}$ is weak* continuous on $R(T_0')$, satisfying (b). Finally, the idempotent mapping J' of $G(T_1)$ on $N(T_1)\times\{0\}$ defined by $J'(e, T_1e) = (e - T^{-1}(T_1e), 0)$ is obviously closed and hence continuous. If $M = (I-J')G(T_1)$, M is a closed complement for $N(T_1)\times\{0\}$ in $G(T_1)$.

In the proof of the sufficiency of conditions (a), (b), and (c) we proceed as follows. Since $N(T_1)\times\{0\}$ has a closed complement M in $G(T_1)$ by (c), there exists a continuous idempotent mapping J of $G(T_1)$ on M such that $J(e,0)=0$ for $e\in N(T_1)$. Since T_1^* is the weak* closure of T_0', and since $(T_0')^{-1}$ exists and is weak* continuous, it follows that $(T_1^*)^{-1}$ exists and is weak* continuous. It follows from Theorem 1.2 of Section 1 that $R(T_1) = F$. Let $G(T)$, the graph of $T, = M$. M is a closed subspace of $G(T_1)$, (and hence of $E\times F$) which includes $G(T_0)$ and is contained in $G(T_1)$. If $e\in D(T_1)$, $J(e, T_1e) = (J_1e, T_1e)$ since J maps $N(T_1)\times\{0\}$ into 0. Hence, $J_1e\in D(T)$, $T(J_1e) = T_1(e)$, and $R(T) = F$. Since F is a Fréchet space and T^{-1} is a closed linear mapping whose domain is the whole of F and maps F into E, T^{-1} is continuous and T is a solvable operator for the pair (T_0, T_0').

Proof of Theorem 3.2: The proof is identical with that of Theorem 3.1 except that with the replacement of (b) by (b)', it no longer follows immediately as before that $(T_1^*)^{-1}$ is weak* continuous. But in the present case, since $E\times F$ is reflexive because E and F are reflexive, the weak* topology on $F^*\times E^*$ coincides with the weak topology. Let G be the graph of T_0' in $F^*\times E^*$. Since G is weakly closed and hence weak* closed, G is the graph of $(T_1)^*$. By the strong continuity of $(T_0')^{-1}$, it follows that the range of T_1^* is closed and T_1^* has a continuous inverse. Hence by Theorem 1.1 of Section 1, $R(T_1) = F$, and the proof proceeds as before.

Remark: Condition (a) of Theorems (3.1) and (3.2) is not used in the sufficiency part of the proof. Indeed, the argument given shows that (b) and (c) together imply (a). The unnecessary condition (a) was introduced to make possible a more explicit comparison with Corollary 3.3, a result for Hilbert spaces essentially due to VISIK [28].

Theorem 3.3: Let E and F be Fréchet spaces. Then a sufficient condition for the pair (T_0, T_0') to have a solvable operator is that (a) and (b) should hold together with the following:

(c)′ $N(T_1)$ has a closed complement in E, while $cl(R(T_0))$ has a closed complement in F.

If E and F are reflexive Banach spaces, the combination of conditions (a), (b)′, and (c)′ is a sufficient condition for the existence of a solvable operator T.

Corollary 3.3: If E and F are Hilbert spaces, a necessary and sufficient condition that the pair (T_0, T_0') should have a solvable operator is that T_0^{-1} and $(T_0')^{-1}$ should be bounded mappings on $R(T_0)$ and $R(T_0')$, respectively.

Proof of Theorem 3.3: It follows from conditions (b) and (b)′ as in Theorems 3.1 and 3.2 that $R(T_1) = F$. Since T_0 is closable, we may assume without loss of generality that T_0 is closed, and hence, since T_0^{-1} is continuous by condition (a), that $R(T_0)$ is closed. The second part of condition (c)′ becomes then: $R(T_0)$ has a closed complement in F. Since $N(T_1)$ has a closed complement in E and $R(T_0)$ has a closed complement in F, it follows from the closed graph theorem for F-spaces that there exist continuous idempotent mappings P_1 of E onto $N(T_1)$ and P_2 of F onto $R(T_0)$. Let $M_1 = (I - P_1)(E_1)$, $M_2 = (I - P_2)(F)$. Let f be an element of M_2. Since $R(T_1) = F$, there exists at least one e in $D(T_1)$ such that $T_1 e = f$. If we define $S_0 f$ as $(I - P_1)e$, the definition of S_0 is unambiguous for every f in M_2 since $T_1 e = T_1 e_1$ implies $e - e_1 \in N(T_1)$, i.e. $(I - P)e = (I - P)e_1$. For each f, $S_0 f$ differs from the original e by an element of $N(T_1)$, and hence $S_0 f$ lies in $D(T_1)$ and $T_1 S_0 f = f$. Furthermore, S_0 is a closed mapping from M_2 into E since if $f_n \to f_0$ and $S_0 f_n \to e_0$, since $T_1 S_0 f_n = f_n$ and T_1 is a closed mapping, e_0 lies in $D(T_1)$ and $T e_0 = f_0$. But M_1 is closed, so that e_0, being the limit of a sequence $S_0 f_n$ from M_1, is itself in M_1. Thus $(I - P_1)e_0 = e_0$, and $e_0 = S_0 f_0$. Applying the closed graph theorem to the closed linear mapping S_0 of the F-space M_2 into the F-space E, we see that S_0 is continuous.

Now, for general f in F, let $Sf = S_0(I - P_2)f + T_0^{-1} P_2 f$. S being the sum of two continuous linear mappings is itself continuous. For f in $R(T_0)$, $Sf = T_0^{-1} f$. For every f, Sf lies in $D(T_1)$ and

$$T_1 S f = T_1 S(I - P_2) f + T(T_0^{-1} P_2 f) = (I - P_2 f + P_2 f = f\,.$$

It follows that $T = S^{-1}$ is contained in T_1, contains T_0, and has the continuous inverse S defined on the whole of F. Thus the theorem has been proved.

Proof of Corollary 3.3: The necessity of conditions (a) and (b) follows from Theorem (3.2). Their sufficiency follows from Theorem (3.3) and the fact that if E and F are Hilbert spaces, they have continuous projections on all their closed subspaces, and (c)′ is automatically satisfied.

In order to characterize the solvable operators, we shall use the formalism introduced by Hörmander [14], Chap. 1. We introduce the space Γ of abstract boundary conditions by $\Gamma = G(T_1)/G$, where G is the closure of the graph of T_0, and the identification mapping γ mapping $G(T_1)$ onto Γ. To every operator T such that $T_0 \subseteq T \subseteq T_1$, there corresponds a subset Γ_T of Γ, where

$\Gamma_T = G(T)/G$. Since we consider only closed operators T containing T_0, all the Γ_T we shall obtain will be closed subspaces of the linear space Γ. On the other hand, if Γ_1 is a closed subspace of Γ, for the operator T whose graph is $\gamma^{-1}(\Gamma_1)$, $\Gamma_T = \Gamma_1$. Thus there is a one-to-one correspondence between closed subspaces of Γ and closed operators T between T_0 and T_1.

The question which is now posed is the following: To give a necessary and sufficient condition on Γ_T in order that T should be a solvable operator for the pair (T_0, T_0'). The answer which is given in Theorem (3.4) is a generalization of a result of HÖRMANDER [14] for the case of Hilbert spaces (which, in turn, generalizes results of VISIK [28] in a slightly different formulation).

Before stating and proving Theorem (3.4), we observe the following fact which is of interest in connection with that theorem:

Lemma 3.1: Let E and F be Fréchet spaces, and let the subspace N of Γ be defined as $\gamma(N(T_1)\times\{0\})$. Then if T^{-1} is a continuous mapping of $R(T_0)$ into F, N is a closed subspace of Γ and γ maps $N(T_1)\times\{0\}$ bicontinuously onto N.

Proof: Since T_0^{-1} is well-defined, it follows that $N(T_1)\times\{0\}$ has no points in common in $G(T_1)$ with the closed subspace $G = c\,l(G(T_0))$. It follows that γ is a one-to-one mapping of $N(T_1)\times\{0\}$ onto N. Further, γ is continuous. To prove that γ is bi-continuous, it suffices by the open mapping theorem for F-spaces to prove that N is a closed subspace of Γ. For this latter purpose, however, we observe that N is closed in Γ iff $\gamma^{-1}(N) = (N(T_1)\times\{0\}) + G$ is closed in $G(T_1)$. Suppose, then, that we have a sequence $g_n = (r_n, 0) + (f_n, T_0 f_n)$ converging to $(g_0, T_1 g_0)$ in $G(T_1)$ with $r_n \in N(T_1)$, $f_n \in D(T_0)$. Without any loss of generality, by taking the closure of T_0, we may assume that G is actually the graph of T_0 and $R(T_0)$ is closed. Since the sequence $\{g_n\}$ converges in $E\times F$, it follows that $T_0 f_n$ converges to an element $T_0 f_0$ of the closed subspace $R(T_0)$ of F. Since T_0^{-1} is continuous as a mapping of $R(T_0)$ into E, it follows that f_n converges to f_0 in E. But then r_n converges in E to an element r_0 of E, which must lie in $N(T_1)$, since this latter subspace is closed. In particular, the limit of g_n as $n\to\infty$ must be $(r_0, 0) + (f_0, T_0 f_0)$, which implies that $\gamma(N(T_1)\times\{0\})$ is closed in Γ.

Theorem 3.4: Let E and F be Fréchet spaces, T_0 and T_0' formally adjoint mappings from E into F and F^* into E^*, respectively, such that T_0^{-1} is continuous and $(T_0')^{-1}$ is weak* continuous. Let $N = \gamma(N(T_1)\times\{0\})$. Then for a closed subspace Γ_0 of Γ; $\gamma^{-1}(\Gamma_0)$ is the graph of a solvable operator for the pair (T_0, T_0') iff Γ is the direct sum of Γ_0 and N. If E and F are reflexive Banach spaces, the weak* continuity of $(T_0')^{-1}$ may be replaced by continuity.

Proof of Theorem 3.4: The conclusion of the present theorem follows from the proof of Theorem 3.1 (or Theorem (3.2) in the case where E and F are reflexive Banach spaces) if we take $M = \gamma^{-1}(\Gamma_0)$. The result follows from the remark that $G(T_1)$ is the direct sum of M and $N(T_1)\times\{0\}$ iff Γ is the direct sum of Γ_0 and N.

Our discussion so far in this section has been restricted to criteria for the existence of solvable operators T. Corresponding results may be obtained for

the question of the existence of completely solvable operators. We shall restrict ourselves to the following simplest and most easily verified sufficient condition:

Theorem 3.5: Let E and F be reflexive Banach spaces, (T_0, T_0') a formally adjoint pair of operators. Then a sufficient condition for the existence of a completely solvable operator for the pair (T_0, T_0') is that all the three following conditions should be satisfied:

(i) T_0^{-1} exists and maps bounded sets of $R(T_0)$ into compact subsets of E.

(ii) $(T_0')^{-1}$ exists and maps bounded subsets of $R(T_0')$ into compact subsets of F^*.

(iii) There exist closed subsets M_1 of E and M_2 of F such that E is the direct sum of M_1 and $N(T_1)$, while F is the direct sum of M_2 and $cl(R(T_0))$.

Proof of Theorem 3.5: We may assume without loss of generality that T_0 and T_0' are closed and hence from (i) and (ii) that $R(T_0)$ is closed in F and $R(T_0')$ is closed in E^*. By assumption (iii) and the closed graph theorem, there exist bounded idempotent mappings P_1 of E on $N(T_1)$ and P_2 of F on $R(T_0)$. It is easy to verify that the mapping $(I - P_0^*)$ is an idempotent mapping of E^* onto $R(T_0')$ (the latter, since it is closed, being the annihilator of $N(T_1)$).

Consider the mapping S_1^* of E^* into F^* defined by $S_1^* = (T_0')^{-1}(I - P_1^*)$. If G_1 is the graph of S_1^* in $E^* \times F^*$, G_1 contains the graph of $(T_0')^{-1}$. Let j be the canonical mapping of $E^* \times F^*$ into $F^* \times E^*$ obtained by reversing the order of the two factors. Since $j(G_1)$ contains the graph of T_0', $(j(G_1))^0$ in $E \times F$ is contained in $(G(T_0'))^0 = G(T_1)$. But $(j(G_1))^0 = j(G(S_1))$, where S_1 is the adjoint of the previously defined operator S_1^*. Thus S_1 maps F into E, and for each f in F, $S_1 f$ lies in $D(T_1)$ and $T_1(S_1 f) = f$. In addition, S_1, being the adjoint of the compact mapping S_1^*, is a compact mapping of F into E.

We now define a mapping S of F into E, by

$$Sf = T_0^{-1} P_2 f + (I - P_1)\, S_1 (I - P_2) f\,.$$

S, being the sum of two compact linear mappings of F into E, is itself a compact linear mapping of F into E. For each f in F, Sf lies in $D(T_1)$, and $T_1 S f = T_0 T_0^{-1} P_2 f + T_1 S_1 (I - P_2) f = P_2 f + (I - P_2) f = f$. If we set T equal to T_1 restricted to the range of S, T is a completely solvable operator for the pair (T_0, T_0').

We conclude this section by specializing the problem to a situation in which the existence of a canonically defined solvable operator may be assured on the basis of assumptions made upon T_0 alone. The construction is due to Friedrichs in the case of symmetric operators in Hilbert-space, and has been applied to the non-symmetric case in various studies of boundary-value problems for elliptic and parabolic equations by a large number of writers (e.g. [3], [4], [5], [6], [7], [12], [18], [19], [20], [21], [27], [29], [30]).

We shall no longer assume that E and F are necessarily Fréchet spaces.

We begin with an operator T_0 having a dense domain in E and range in F. Suppose that there exists a linear mapping J of $D(T_0)$ onto a dense subset

of F^* such that

$$(3.1)\qquad Re\,\langle T_0e, Je\rangle > 0\,,\quad e \in D(T_0)\,.$$

We form the pre-Hilbert space structure on $D(T_0)$ with the inner product $[e, e_1] = \frac{1}{2}\{\langle T_0e, Je_1\rangle + \overline{\langle T_0e_1, Je\rangle}\}$. The corresponding norm is defined by $\|e\|_H^2 = [e, e]$.

We complete $D(T_0)$ to the Hilbert space H with respect to the H-norm which we have just defined. We shall assume that the Hilbert space H can be continuously imbedded in the space E. If E is sequentially complete the existence of a continuous extension of the identity map of $D(T_0)$ into E is equivalent to the following: For each semi-norm $\|\cdot\|_\beta$ on E corresponding to a convex neighborhood of the origin in E, there exists a constant $c_\beta > 0$, such that

$$(3.2)\qquad \|e\|_H^2 = Re\,\langle T_0e, Je\rangle \geqq c_\beta^2\,\|e\|_\beta^2\,,\quad e \in D(T_0)\,.$$

Indeed, (3.2) implies that a Cauchy sequence in H is a Cauchy sequence in the natural uniformity of E, and from the sequential completeness of E, it follows that there is an uniquely defined continuous mapping j of H into E such that j is the identity mapping on $D(T_0)$. We shall assume also that j is one-to-one. This will follow, for example, if J is continuous as a mapping from the subspace $D(T_0)$ into F^* and if $R(J) \subset D(T_0')$. For in that case, suppose that $\{e_n\}$ is a Cauchy sequence in H which converges to 0 in E, and let e_0 be an arbitrary element of $D(T_0)$, h the limit of e_n in H. Then we have

$$[h, e_0] = \lim_n [e_n, e_0] = \frac{1}{2}\lim_n\{\langle T_0e_n, Je_0\rangle + \overline{\langle T_0e_0, Je_n\rangle}\}$$
$$= \frac{1}{2}\lim_n\{\langle e_n, T_0'Je_0\rangle + \overline{\langle T_0e_0, Je_n\rangle}\} = 0\,,$$

and since $D(T_0)$ is dense in H, h must $= 0$.

In any case, we shall assume that j is one-to-one and that J is a continuous mapping from $D(T_0)$ considered as a subspace of H into F^*. It follows that J can be extended by continuity to a continuous linear mapping of all of H into F^*.

For each f in F, we have by the last assumption that $\langle f, Je\rangle$ is a bounded conjugate linear functional of e in H. By the classical theorem of Fréchet and Riesz on the representation of linear functionals on a Hilbert space, there exists an unique element Gf of H such that

$$(3.3)\qquad [Gf, e_1] = \langle f, Je_1\rangle$$

for all e_1 in H, since J being continuous maps bounded sets in H into bounded sets of F^*. Furthermore, G is a continuous linear map of F into H. By the same argument, for every element e in $D(T_0)$, the linear product $\langle Te, Je_1\rangle$ is a continuous conjugate linear functional of e_1 in H and therefore may be represented uniquely in the form

$$(3.4)\qquad [S_0e, e_1] = \langle Te, Je_1\rangle$$

for all e_1 in H. For the operator S_0 in H as thus defined we have

$$(3.5)\qquad Re\,[S_0e, e] = Re\langle Te, Je\rangle \leqq \|e\|_H^2\,.$$

We shall assume finally that one of the two following conditions holds for T_0:

(I) $|\langle Te, Je_1\rangle| \leq c\|e\|_H \|e_1\|_H$ for some constant c and all e, e_1 in $D(T_0)$;

(II) If for a given h in H, we have $\langle Te, Jh\rangle = 0$ for all e in $D(T_0)$, then $h = 0$.

It follows from Theorem 1.5 of Section 1 that S_0 has a closure S in H, that the range of S is closed and is the closure of the range of S_0 and that S^{-1} is a bounded mapping on $R(S)$. If (II) holds, it follows that $R(S_0)$ is dense, since its orthogonal complement in S is just $\{0\}$, and therefore $R(S) = H$. If (I) holds, S_0 and hence S are bounded linear mappings of H, and it follows from (3.5) and Theorem 1.7 of Section 1 that $R(S) = H$. In both cases, we are assured that $R(S) = H$.

Definition 3.4: The generalized Friedrichs extension T of T_0 with respect to E, F, and J is defined as follows: $D(T) = \{e : e = j(h)$ for $h \in D(S)$, with $Sh = Gf$ for some f in $F\}$, $Te = f$. More succinctly, $T = j^{-1}SG^{-1}$.

Theorem 3.6: Let E and F be two locally convex linear spaces, T_0 a densely defined linear mapping from E to F, T_0' a weak* densely defined linear mapping from F^* to E^* which is formally adjoint to T_0. Suppose that there exists a linear mapping J of $D(T_0)$ into F^* such that $Re\langle T_0 e, Je\rangle > 0$ for all e in $D(T_0)$, that the Hilbert space H obtained by completing $D(T_0)$ with respect to the norm $\|e\|_H^2 = Re\langle T_0 e, Je\rangle$ is continuously imbeddable in E, and that J is a continuous linear mapping from $D(T_0)$ in the H-norm into F. Suppose further that one of the two following assumptions is satisfied:

(I) There exists a constant $c > 0$ such that $|\langle Te, e_1\rangle| \leq c\|e\|_H\|e_1\|_H$;

(II) If for $h \in H$, $\langle Te, Jh\rangle = 0$ for all e in $D(T_0)$, then $h = 0$.

Then the generalized Friedrichs extension T of T_0 is a closed mapping of $D(T)$ onto F such that $T_0 \subseteq T$ and having a continuous inverse.

If in addition, the image of H under J contains $D(T_0')$, then T is a solvable operator for the pair (T_0, T_0).

Proof of Theorem 3.6: We first remark from Definition 3.4 that since j and G are one-to-one mappings, T is well defined. The inverse T^{-1} of T is precisely $jS^{-1}G$, and since S^{-1} is everywhere defined and continuous as a mapping of H by our assumptions (I) and (II), T^{-1} is everywhere defined and continuous as a mapping from F into E. If $e \in D(T_0)$, then $\langle T_0 e, Je_1\rangle = [Se, e_1] = [Ge, e_1]$. It follows that $e = j^{-1}SG^{-1}(T_0 e)$, i.e. $e \in D(T)$ and $T_0 e = Te$. Thus, $T_0 \subseteq T$.

Suppose now that $J(H)$ contains $D(T_0')$. For every e in $D(T)$, there exists a sequence $\{e_k\}$ in $D(T_0)$ such that $e_k \to e$ in H, $Se_k \to Se$ in H. In particular, it follows that for h in H, $[S_0 e_k, h] = \langle T_0 e_k, Jh\rangle \to [Se, h] = \langle Te, Jh\rangle$. But for each k, $\langle T_0 e_k, f^*\rangle = \langle e_k, T_0' f^*\rangle$ for each f^* in $D(T_0')$, while the convergence of e_k in H implies the convergence of the sequence to e in E, which implies in turn that $\langle e_k, T_0' f^*\rangle \to \langle e, T_0' f^*\rangle$. Since $J(H)$ contains $D(T_0')$, we have on the one hand $\lim_k \langle T_0 e_k, f^*\rangle = \langle Te, f^*\rangle$, and on the other hand $\lim_k \langle T_0 e_k, f^*\rangle = \langle e, T_0' f^*\rangle$. Thus for each e in $D(T)$ and each f^* in $D(T_0')$, $\langle Te, f^*\rangle = \langle e, T_0' f^*\rangle$. It follows, since T_1 is the restricted adjoint of T_0', that $T \subseteq T_1$, and hence that T is a solvable operator for the pair (T_0, T_0').

Bibliography

[1] Banach, S.: Théorie des Opérations Linéaires. Warsaw 1932. — [2] Bourbaki, N.: Espaces Vectoriels Topologiques. Paris 1953. — [3] Browder, F. E.: The Dirichlet problem for linear elliptic equations of arbitrary order. Proc. nat. Acad. Sci. (Wash.) **38**, 230—235, 741—747 (1952). — [4] Browder, F. E.: Linear parabolic equations of arbitrary order: general boundary value problems for elliptic equations. Proc. nat. Acad. Sci. (Wash.) **39**, 185—190 (1953). — [5] Browder, F. E.: Strongly elliptic systems of differential equations. Ann. of Math. Study No. 33, 15—51 (1954). — [6] Browder, F. E.: On the regularity properties of solutions of elliptic differential equations. Comm. Pure appl. Math. **9**, 351—361 (1956). — [7] Browder, F. E.: Parabolic systems of differential equations with time dependent coefficients. Proc. nat. Acad. Sci. (Wash.) **42**, 914—917 (1956). — [8] Collins, H. S.: Completeness and compactness in linear topological spaces. Trans. Amer. math. Soc. **74**, 256—281 (1955). — [9] Dieudonné, J., and L. Schwartz: La dualité dans les éspaces (F) et (LF). Ann. Inst. Fourier **1**, 61—101 (1949). — [10] Fichera, G.: Alcuni recenti sviluppi della teoria dei problemi al contorno per le equazioni alle derivate parziali lineari. Atti Congress sulle equazioni a derivate parz. di Trieste (1954), pp. 174—227. Roma: Ediz. Cremonese 1955. — [11] Fichera, G.: Su un principio di dualita per talune formule di maggiorazione relativa alle equazioni differenziali. R. C. Accad. Naz. Lincei, S VIII, **19**, 411—418 (1955). — [12] Garding, L.: Dirichlet's problem for linear elliptic partial differential equations. Math. scand. **1**, 55—72 (1953). — [13] Hille, E., and R. S. Phillips: Functional Analysis and Sèmi-Groups, (Revised Edition). Amer. Math. Soc. Colloquium Publ. **31**, Providence (1957). — [14] Hörmander, L.: On the theory of general partial differential operators. Acta math. **94**, 161—248 (1955). — [15] Kelley, J. L.: The closed graph theorem and the Krein-Smulian theorem. Techn. Rep. No. 20, O.N.R. Contract Nonr. 222 (37). University of California at Berkeley, January 1958. — [16] Köthe, G.: Zur Theorie der kompakten Operatoren in lokalkonvexen Räumen. Portugalia Math. **13**, 97—104 (1954). — [17] Lax, P., and A. N. Milgram: Parabolic equations. Ann. of Math. Study No. **33**, 167—190 (1954). — [18] Leray, J.: Hyperbolic Equations. Institute for Advanced Study Notes, Princeton 1954. — [18] Lions, J. L.: Problèmes aux limites en théorie des distributions. Acta math. **94**, 13—153 (1955). — [19] Lions, J. L.: Sur les problèmes aux limites du type derivée oblique. Ann. of Math. **64**, 207—239 (1956). — [20] Lions, J. L.: Sur les problèmes mixtes pour certains systems paraboliques dans les ouverts non-cylindriques. To appear in the Annales de l'Institut Fourier. — [21] Nirenberg, L.: Remarks on strongly elliptic partial differential equations. Communications Pure appl. Math. **8**, 643—674 (1955). — [22] Pták, V.: On complete topological linear spaces. Cehoslovack Mat. Z. **1953**, 301—364. — [23] Robertson, A. P., and W. Robertson: On the closed graph theorem. Proc. Glasgow Math. Ass. **3**, 9—13 (1956). — [24] Schwartz, L.: Théorie des distributions. Paris 1951. — [25] Schwartz, L.: Homomorphismes et applications complètement continues. C. R. Acad. Sci. (Paris) **1953**, 2472—2473. — [26] Serre, J. P.: Exposé XVI, Séminaire de H. Cartan, Paris 1954. — [27] Visik, M. I.: On strongly elliptic systems of differential equations. Mat. Sbornik **29**, 615—676 (1951). — [28] Visik, M. I.: On general boundary problems for elliptic differential equations. Trudi Moscow Math. Soc. **1**, 187—246 (1952). — [29] Visik, M. I.: The Cauchy problem for equations with operator coefficients, mixed boundary problems for systems of differential, equations, and approximation methods for their solution. Mat. Sbornik **39** (81), 50—148 (1956). — [30] Visik, M. I., and O. A. Ladyzenskaya: Boundary value problems for partial differential equations and certain classes of operator equations. Uspekhi Mat. Nauk. **11**, 41—97 (1956). — [31] Yood, B.: Properties of linear transformations preserved under addition of a completely continuous transformation. Duke math. J. **18**, 599—612 (1951).

(Eingegangen am 12. Februar 1959)

Groot de J.
Math. Annalen 138, 80—102 (1959)

Groups represented by homeomorphism groups I.

By

J. de Groot in Amsterdam

Notations

Group stands for *abstract* group.

$|G|$, $|S|$ denote the order of a group G or, in general, the cardinal of a set S.

$A(G)$, $A(T)$ denotes the automorphism group of a group G (or more generally of some algebraic structure) or the autohomeomorphism group (group of *all* topological transformations) of a topological space T.

$T(G)$, $T(T)$ denote subgroups of $A(G)$ and $A(T)$ respectively.

$\emptyset$ denotes the empty set. Bold face letters $\mathbf{k}, \mathbf{m}, \mathbf{n}, \ldots$ stand for infinite cardinals, $\aleph_0$ and $\aleph$ are the cardinals of the sets of natural and real numbers respectively.

$S_{\mathbf{m}}$ denotes the (full) symmetric group of a set M of cardinal $|M| = \mathbf{m}$.

Contents

1. Introduction

The present paper arose from the following problem.

Let a group G be given. To what extent can one find some satisfactory topological space T such that G is isomorphic to $A(T)$: $G \simeq A(T)$?

R. J. Wille and the author have already shown [13] that if G is *countable*, there always exists a Peanocurve P (1-dimensional, connected, locally connected continuum) for which $G \simeq A(P)$. One of the results we shall establish here, is an affirmative answer to the general question: for *every* group G one can find a complete, connected, locally connected *metric* space M of any *positive* dimension such that $G \simeq A(M)$ (see theorem 7). A curious purely algebraic corollary to this result is theorem 10: *Every* group is the automorphism group of some commutative ring [12].

Our metric space M is constructed in several steps. Given G, we construct the Cayley-graph G^* of G (section 6). G^* is a coloured, directed graph such that

$A(G^*) \simeq G$. From G^* we then construct a graph $\mathfrak{G}$ such that $A(\mathfrak{G}) \simeq G$. The last step, which furnishes the required M, is to replace each edge of the graph $\mathfrak{G}$ by a suitable rigid space and to introduce a metric in the resulting set.

A topologically *rigid* space is a space T for which $A(T) = \{e\}$ (identity). Rigid spaces and also rigid algebraic structures, e. g. graphs, play an important part in our investigations. Rigid spaces were introduced in [13] by effective methods[1]). Here we shall use the axiom of choise freely, and the results will be far stronger. We prove e. g. (section 3) the existence of subsets of the line which are rigid for topological maps into, i. e. which admit no topological transformations into themselves (except the identity) and which are rigid for continuous maps onto. In the plane we find connected, locally connected subsets which are rigid for continuous maps into, i. e. which do not admit any continuous maps into themselves (except the identity map and the map onto a single point).

The rigid spaces, as mentioned, provide us with a set of useful examples and counterexamples. They themselves give a final answer to certain inquiries, initiated by the Polish mathematicians (cf. Kuratowski [18], Chapter III, § 31; see also [11]).

The results on rigidity are obtained by introducing the concept of displacement. A set-theoretical *displacement* is, roughly speaking, concerned with the number of points which are moved into other points under some given map. A *continuous displacement* is defined as a continuous map which displaces at least $\varkappa$ points into $\varkappa$ other points. So, in many ways, this concept is contrasting to the concept of a fixed point. Rigid spaces are now constructed by (i) recognizing continuous maps in many cases as continuous displacements and then (ii) inventing means to forbid them; but forbidding them means rigidity. It is believed that there will be other applications of the displacement concept.

There is, as usual, a striking difference between the results for positive dimension and those for dimension zero. This is also true for our initial problem and so, in general, the question arises, what can be said about the groups $A(N)$ and $T(N)$, if N is a 0-dimensional space. Using the concept of equivalence (§ 4), we first obtain the simple (but to the author surprising) result (see § 4, formula 4.1), that a group is a group of autohomeomorphisms $T(D_\mathfrak{m})$ of the generalized discontinuum $D_\mathfrak{m}$ (topological product of $\mathfrak{m}$ pairs of points), if and *only if* it is some group of permutations of $\mathfrak{m}$ objects (subgroup of $S_\mathfrak{m}$).

From this follows e. g. that every abelian group of order $\leqq 2^\mathfrak{m}$ is a group $T(D_\mathfrak{m})$. This, by no means, indicates the general situation for an $A(H)$ with 0-dimensional, bicompact Hausdorff H. For such H there are two important pathologies: (i) rigid H, for which $A(H) = \{e\}$, (ii) H, for which $A(H)$ consists of an infinite number of elements, each of order $\leqq 2$.

[1]) A natural procedure to construct rigid spaces is as follows. Take care that the space contains an everywhere dense (preferably countable) subset such that each point of this set has a property (e. g. its order, if the space is a curve) unique in the entire space. So, a rigid curve is easily constructed. Added in proof: cf. also A. S. Besicovitch, totally heterogeneous continua, Proc. Cambridge Phil. Soc. **41**, 96—103, (1945).

The notions of equivalence and rigidity are further exploited to obtain information on the $A(N)$ of 0-dimensional, separable, metrizable spaces N (§ 5) and on the $A(B)$ of Boolean rings B (§ 8.1). Furthermore, the center of the group $A(N)$ is determined if N is an arbitrary 0-dimensional set of real numbers (theorem 4). In particular, it follows that, usually, a given group cannot be represented by an $A(N)$: the order of $A(N)$, if N is 0-dimensional, dense in itself and metrizable, is e. g. either 1 or $\geq \aleph$.

Summarizing, the following situation is obtained.

Every group may be the auto(homeo)morphism group of a graph, a distributive lattice (BIRKHOFF), a commutative ring or a bicompact, connected Hausdorff space. On the other hand, a group is not, in general, the auto(homeo)-morphism group of a group, a 0-dimensional bicompact Hausdorff space or a complemented, distributive lattice. For this failure there may be two reasons: either there is no rigidity (e. g. every not too small abstract group has proper automorphisms), or proper rigidity may occur (as in the case of Boolean algebras of continuum order), but if the structure fails to be entirely rigid, the existence of one proper auto(homeo)morphism may lead to a shower of (infinitely) many others.

Knowledge concerning the possible $A(G)$ of topological groups G is still scarce. A forthcoming paper by H. DE VRIES [23] gives useful information, but up to this moment it is still unknown whether there exist rigid topological groups.

2. Displacements

Let there be given an abstract set N and a map f of a subset S of N into N. The set fS is also denoted by fN (no confusion will arise), and more generally fM denotes $f(S \cap M)$.

Definition: f is a *displacement* of order $\mathfrak{m}$, if there is a subset V of N such that

(2.1) $$V \cap fV = \theta, \ |fV| = \mathfrak{m} \ \ (\mathfrak{m} \text{ maximal}),$$

i. e. there is no greater cardinal than $\mathfrak{m}$ for which a subset $V \subset N$ exists such that V does not intersect its image fV and $|fV|$ equals this cardinal. Observe that $|V| \geq \mathfrak{m}$. If $W \cap fW = \theta$, $|fW| = \mathfrak{k}$, the set fW is called a *displaced set* of order $\mathfrak{k}$.

For every N and map f there exists some cardinal $\mathfrak{m}$, such that f is a displacement of order $\mathfrak{m}$. (Clearly $\mathfrak{m} \leq |N|$). The non-trivial proof of this assertion will not be given, since it is nowhere used in the sequel.

Observe furthermore the trivial fact that if f is extended in some way over some subset of N, the extended f is a displacement of at least the same order as f. So, if f has the maximal order, i. e. the order of f equals $|N|$, then the extended displacement has the same order.

Lemma 1: Let N be a set, $|N| = \mathfrak{m}$, and let $\{f_\beta\}$ be a system of $\mathfrak{m}$ displacements f_β (the indices running over a set of potency $\mathfrak{m}$), each of order $\mathfrak{m}$.

Then there is a family $\{F_\gamma\}$ of $2^{\mathfrak{m}}$ subsets $F_\gamma \subset N$, such that for every triple of indices β, γ, γ':

$$|F_\gamma \setminus F_{\gamma'}| = \mathfrak{m} \quad (\gamma \neq \gamma')\,, \tag{2.2}$$

$$|f_\beta F_\gamma \setminus F_{\gamma'}| = \mathfrak{m}\,. \tag{2.3}$$

Moreover, if $\{K_\beta\}$ is a family of $\mathfrak{m}$ subsets $K_\beta \subset N$, $|K_\beta| = \mathfrak{m}$ for every β, then we can require for every pair β, γ:

$$|F_\gamma \cap K_\beta| = \mathfrak{m}, \; |(N \setminus F_\gamma) \cap K_\beta| = \mathfrak{m}\,. \tag{2.4}$$

Proof. Well-order all pairs f_α, K_α:

$$f_1, K_1, f_2, K_2, \ldots, f_\omega, K_\omega, \ldots, f_\alpha, K_\alpha, \ldots \; (\alpha < \omega_{\mathfrak{m}})\,, \tag{2.5}$$

where $\omega_{\mathfrak{m}}$ is the smallest ordinal of potency $\mathfrak{m}$, such that every pair occurs $\mathfrak{m}$ times in this transfinite sequence (this is possible, since $\mathfrak{m}^2 = \mathfrak{m}$).

We shall select a subset $P \subset N$. Consider the first pair f_1, K_1. There are distinct elements p_1 and p_1' which differ from $f_1 p_1 = q_1$ and $f_1 p_1' = q_1' \neq q_1$, since f_1 is a displacement of order $\mathfrak{m}$. Take moreover distinct elements $r_1, s_1 \in K_1$, distinct from p_1, p_1', q_1, q_1'. Consider the sets $\{p_1, p_1', r_1\}$ and $\{q_1, q_1', s_1\}$. We use transfinite induction with respect to the index α in (2.5). Assume disjoint sets

$$\{p_\xi, p_\xi', r_\xi\}_{\xi < \alpha}, \{q_\xi, q_\xi', s_\xi\}_{\xi < \alpha}\,, \tag{2.6}$$

consisting of pairwise different elements have been defined, such that

$$f_\xi p_\xi = q_\xi, \, f_\xi p_\xi' = q_\xi'\,, \quad r_\xi, s_\xi \in K_\xi\,. \tag{2.7}$$

Now we select distinct elements $p_\alpha, p_\alpha', q_\alpha, q_\alpha', r_\alpha, s_\alpha$ not occurring in the sets of (2.6) such that

$$f_\alpha p_\alpha = q_\alpha, f_\alpha p_\alpha' = q_\alpha'\,, \quad r_\alpha, s_\alpha \in K_\alpha\,.$$

This is possible, since the sets (2.6) contain fewer than $\mathfrak{m}$ elements, while f_α is a displacement of order $\mathfrak{m}$ and $|K_\alpha| = \mathfrak{m}$. In this way sets corresponding to (2.6) have been defined, but now for all $\xi \leq \alpha$. Now we can define

$$P = \{p_\xi, p_\xi', r_\xi\}_{\xi < \omega_{\mathfrak{m}}}\,, \quad Q = \{q_\xi, q_\xi', s_\xi\}_{\xi < \omega_{\mathfrak{m}}}\,.$$

Then (2.7) holds for all $\xi < \omega_{\mathfrak{m}}$, $P \cap Q = 0$, and all symbols occurring in the definition of P and Q denote different elements.

For every index β, $f_\beta P$ contains $\mathfrak{m}$ elements, which do not belong to P. Indeed an $f_\beta \in \{f_\beta\}$ occurs $\mathfrak{m}$ times in (2.5). So there are $\mathfrak{m}$ indices α for which

$$f_\beta p_\alpha = f_\alpha p_\alpha = q_\alpha \notin P\,,$$

and this set $\{q_\alpha\}$ has potency $\mathfrak{m}$. Furthermore, P as well as Q contain $\mathfrak{m}$ elements from each K_β, i. e. the sets $\{r_\alpha\}$ and $\{s_\alpha\}$ respectively.

Finally we define the family $\{F_\gamma\}$. Consider the set $\{p_\xi'\}_{\xi < \omega_{\mathfrak{m}}}$ of potency $\mathfrak{m}$ and determine $2^{\mathfrak{m}}$ subsets T_γ of it, such that $|T_\gamma \setminus T_{\gamma'}| = \mathfrak{m}$ for every pair of distinct indices γ, γ' (cf. [18], p. 330). Now define

$$F_\gamma = \{p_\xi\}_{\xi < \omega_{\mathfrak{m}}} \cup \{r_\xi\}_{\xi < \omega_{\mathfrak{m}}} \cup T_\gamma\,.$$

In view of the properties of P and Q mentioned, it is easy to see that $\{F_\gamma\}$ is a family which satisfies the requirements of our lemma.

This lemma admits several applications in general topology. We need the following

Definition: If N is a topological space, f is called a *continuous displacement*, if f is a continuous map of a subset of N into N, and f is a displacement of order $\varkappa$.

Theorem 1. *Let M be a complete separable metric space with $|M| = \varkappa$. Then there exists a family $\{F_\gamma\}$ of $2^\varkappa$ subsets $F_\gamma \subset M$ satisfying the relations*

$$(2.8) \qquad |F_\gamma \backslash F_{\gamma'}| = \varkappa \,, \quad \textit{for all } \gamma, \gamma' \textit{ with } \gamma \neq \gamma'$$

such that no F_γ admits any continuous displacement into itself or into another $F_{\gamma'}$. Moreover, if $\{K_\beta\}$ is a family of $\varkappa$ subsets of M, each of potency $\varkappa$, then for every pair β, γ we may require

$$(2.9) \qquad |F_\gamma \cap K_\beta| = \varkappa \,, \quad |(M \backslash F_\gamma) \cap K_\beta| = \varkappa \,.$$

Proof. There exist only $\varkappa$ many G_δ-subsets of M and a fixed subset of M admits only $\varkappa$ many continuous maps into M, so certainly at most $\varkappa$ different continuous displacements. Let f_β be a continuous displacement, its domain being a G_δ-subset of M and its range contained in M. Hence the family $\{f_\beta\}$ has at most continuous potency. If there does not exist a continuous displacement in M, the theorem is trivial. Thus we may assume that the family $\{f_\beta\}$ has continuous potency in any case, counting a given continuous displacement $\varkappa$ times (if necessary).

Now we apply lemma 1 with $N = M$, $\mathfrak{m} = \varkappa$, $\{f_\beta\} = \{f_\beta\}$. Hence there is a family $\{F_\gamma\}$ of $2^\varkappa$ subsets $F_\gamma \subset M$ satisfying (2.8) and (2.9). We still have to prove that F_γ does not admit *any* continuous displacement φ into itself or into another $F_{\gamma'}$. The map φ can be extended (as is well known) to a continuous map $\tilde{\varphi}$ of a G_δ-set $\bar{F}_\gamma \subset F_\gamma$ into M. $\tilde{\varphi}$ is also a continuous displacement, hence $\tilde{\varphi} = f_\beta$ for some β. So according to (2.3), for every pair γ, γ'

$$f_\beta F_\gamma \backslash F_{\gamma'} = \theta \,.$$

But $f_\beta = \varphi$ on F_γ, so

$$\varphi F_\gamma \backslash F_{\gamma'} \neq \theta \,,$$

which says that φ maps F_γ neither into itself nor into another $F_{\gamma'}$, which concludes the proof.

3. Rigid subsets of the line and the plane

We want to know conditions under which non-trivial continuous or topological maps are continuous displacements.

A map of a set is said to be *trivial* if it is the identity map or if the image-set of the map consists of one element.

Lemma 2: Let P be a separable metric space in which every point is a point of condensation. If $\varphi: P \to P$ is non-trivial and locally topological into P or continuous onto P then φ is a continuous displacement.

Proof. If φ is locally topological and non-trivial, there are points $a, b \in P$ such that $a \neq b$, $\varphi a = b$. This yields disjoint neighbourhoods $U(a)$ and $U(b)$, such that $|U(b)| = \varkappa$ and $\varphi U(a) = U(b)$. So φ is a continuous displacement.

If φ is continuous *onto* and non-trivial, determine a and b as before. Then there is a neighbourhood $V = V(a)$ with $|V| = \varkappa$ and a $U = U(b)$ disjoint from V such that $\varphi V \subset U$, so

$$V \cap \varphi V = \theta \quad \text{hence} \quad \varphi^{-1} V \cap V = \theta .$$

We conclude that V is a displaced set of order $\varkappa$ under φ.

Remark: Observe that the lemma does not hold for non-trivial continuous maps into.

Lemma 3: *Let P be a locally connected subset of a separable metric space M. Every non-trivial continuous map φ of P into M is a continuous displacement, if and only if P remains connected after the removal of at most one point.*

Proof. The condition is necessary. Let us discuss the case where P is connected but becomes not connected after the removal of a point p. Then there is a separation:

$$P \backslash \{p\} = A \cup B .$$

Now the map defined by $A \cup \{p\} \to \{p\}$, and being the identity on B is non-trivial and continuous, but not a continuous displacement.

To show the sufficiency, we suppose *first* $\varphi P \subset P$ and distinguish between the following cases

(i) $\qquad \varphi \neq e$ (identity) on φP ,

(ii) $\qquad \varphi = e$ on φP .

(i) There is a point $q \in \varphi P$, such that $q \neq \varphi q$. Determine disjoint neighbourhoods $U(q)$ and $V(\varphi q)$, relative to the space φP, such that $\varphi U \subset V$. This means that no point of U is mapped onto itself under φ. Since $U \subset \varphi P$, every point of U is image point. From $|U| = \varkappa$ (φP is connected) it follows that U is a displaced set of order $\varkappa$.

(ii) Topological notions will be relative to P. Since φ is non-trivial, there is a point $q \in P \backslash \varphi P$. Since $\varphi = e$ on φP, this set φP is closed and we can determine for every $r \in P \backslash \varphi P$ an open, connected neighbourhood $U(r)$, not intersecting φP. Consider the family of all "finite chains", the elements of which are elements of $\{U(r) | r \in P \backslash \varphi P\}$ connecting some $U(r)$ with $U(q)$. The union of the elements of this family determines an open, connected set S, closed in $P \backslash \varphi P$. $\bar{S}$ has at least two limitpoints contained in φP, for, if $\bar{S} = S \cup \{t\}$, $t \in \varphi P$, then t would be a separating point of P (we use the fact that φP consists of more than one point). Hence there are two points $u, v \in \varphi P$, such that $S' = S \cup \{u, v\}$ is connected and therefore, $\varphi S'$ is connected. Since $\{u, v\} \subset \varphi S'$, this set $\varphi S'$ is non-degenerate and connected. φS is therefore a displaced set of order $\varkappa$ and φ a continuous displacement.

Secondly, we consider the general case $\varphi P \subset M$. We may assume that φP consists of $P \cap \varphi P$ and of less than continuously many limitpoints of this set in M, since otherwise $\varphi P \backslash P$ clearly is a displaced set of order $\varkappa$. — Hence

we may assume that in $P \cap \varphi P$ every point is a condensation point. Now again we distinguish between the cases

(i) $\varphi \neq e$ on $P \cap \varphi P$, (ii) $\varphi = e$ on $P \cap \varphi P$

In both cases the remaining part of the proof proceeds in about the same way as in the previous cases (i) and (ii) (φP being replaced by $P \quad \varphi P$). This proves our lemma.

Remarks: The local connectedness of P is essential in lemma 3, even in the case where $\varphi P \subset P$. Indeed, if we let radiate from a disc a countable number of radii, everywhere dense on the circumference of the disc, and of equal length, we obtain a set P which remains connected after the removal of a finite number of points and for which the radial projection φ on the disc is a displacement retraction of order $\aleph_0$ only.

We notice that this P is not compact. However, we can construct more sophisticated examples of continua which remain connected after the removal of an arbitrary 0-dimensional (or even n-dimensional) set and nevertheless admit continuous maps which are displacements of order $\aleph_0$.

It is possible to give sufficient conditions for compact P in terms of P and $\varphi P \subset P$, without claiming the local connectedness of P, but these conditions are less natural.

Lemma 4: If P is the complement of a totally imperfect set (a set is totally imperfect, if it does not contain a topological image of the discontinuum of Cantor) in E^2, then P is connected, locally connected and retains these properties after the removal of one point. If P and $E^2 \backslash P$ are totally imperfect, both have dimension 1.

The connectedness and local connectedness of P is a wellknown property derived by SIERPIŃSKI. About the same method of proof shows the second part of the lemma. The 1-dimensionality follows from a classical theorem of BROUWER.

Definition: A set is called *rigid* for a certain class of maps into itself, if the set admits no such maps except trivial ones ("trivial" being defined in some natural way).

More precisely, we shall encounter *topologically rigid spaces*, i.e. the topological space S admits no topological transformation except the identity, spaces *rigid under continuous maps onto* (the only continuous map onto equals identity), spaces *rigid under continuous maps into* (the only such maps are maps onto a point or identity), and rigid graphs (the automorphism group of the graph equals identity) in section 6. The *rigidity of the field of real numbers* is well-known (and this is used in section 8).

Topologically rigid spaces have been introduced in [13]. The following theorems are, however, much stronger.

Theorem 2. *There exists a family $\{F_\gamma\}$ of $2^\aleph$ 0-dimensional subsets of the real line, such that no F_γ can be mapped locally topologically into or continuously onto itself or any other $F_{\gamma'}$. If F_γ is mapped into itself, we must exclude trivial maps.*

Theorem 3. *There exists a family $\{F_\gamma\}$ of $2^\aleph$ 1-dimensional, connected and locally connected subsets of the plane, such that no F_γ can be mapped continuously and non-trivially into itself or another $F_{\gamma'}$.*

In particular we have:

Corollary 1: *The real line contains sets of potency $\aleph$ which are rigid under topological maps into and continuous maps onto. The plane contains connected, locally connected sets, which are rigid under continuous maps into.*

This answers a.o. a problem raised in [11], p. 204, and generalizes the theorems given there.

Corollary 2: *There exist connected, locally connected suosets of the plane, such that every continuous map into itself is a retraction.*

Indeed, the map either equals the identity or its image consists of one point.

Proof of theorem 2. We apply theorem 1, where M is the real line, and $\{K_\beta\}$ the system of all uncountable compact subsets of M. The family $\{F_\gamma\}$ is the required system. Indeed, the F_γ and $M\backslash F_\gamma$ become totally imperfect. So every point of an F_γ is condensation point. Hence, applying lemma 2, we see that every locally topological map φ of an F_γ into itself or a continuous map φ onto itself is either trivial or a continuous displacement. According to theorem 1 it cannot be a continuous displacement. We conclude that φ is trivial.

Secondly, take a pair $F_\gamma, F_{\gamma'}$ $(\gamma \neq \gamma')$. In view of (2.8) and the fact that every point of an F_γ is a condensation point, the reader may easily verify that every locally topological map of F_γ into $F_{\gamma'}$ or continuous map of F_γ onto $F_{\gamma'}$ is a continuous displacement. However, this is impossible according to theorem 1. This proves the theorem.

Proof of theorem 3. We apply theorem 1 in the same way as in the preceding proof. M denotes in this case the Euclidean plane. Applying lemma 4, we see that the elements of $\{F_\gamma\}$ are 1-dimensional, connected and locally connected, and moreover, remain connected, after the removal of one point. So we can apply lemma 3, which gives us together with theorem 1 the required proof.

Remarks: The topologically rigid 0-dimensional sets constructed above, are all totally imperfect. However, applying these results, one can easily construct topologically rigid 0-dimensional sets which contain a discontinuum of Cantor.

The constructions above are ineffective. If we use the family of all sub-continua (of the line or plane) instead of the family of all non-countable compacta, we obtain *punctiform* rigid sets instead of totally imperfect ones. Since punctiform sets can be constructed effectively, it seems probable that also topologically rigid subsets of the line can be constructed effectively.

An effective topologically rigid plane Peanocurve has been described in [13]. We briefly mention its construction (omitting proofs), since it will be used in section 7.

Consider a disc D in the plane. Let $\{a_i\}$ be a countable dense subset of the interior of D. We define a sequence of "propellers" in D.

The first is an $(n+1)$-bladed propeller, i.e. an $(n+1)$-bladed curve having as its center a_1, which avoids the boundary of D. We proceed by induction

and suppose that the first $k-1$ propellers have already been defined. Let a'_k be the first member of the sequence $\{a_i\}$ which is in no previously constructed propeller. Then the k-th propeller is $(n+k)$-bladed with center at a'_k and lies inside a circle which misses all previously constructed propellers and the boundary of D. Moreover, we take care that the diameters of the propellers tend to zero. The topologically rigid space is the disc D with the interiors of all the propellers removed, and we shall denote this space by P_n.

4. Equivalent groups

Two groups G and H are said to be *equivalent* if each of them is isomorphic to a subgroup of the other.

This gives a decomposition of, say, the family of all groups $\{G\}$ of given order $|G| = \mathfrak{m}$ into disjoint *equivalence classes* εG. Two elements G and H of the same equivalence class have an important property in common, as can be easily seen: *if $\varepsilon G = \varepsilon H$, then G and H have the same systems of subgroups*, i.e. if a group S occurs a number of times (isomorphically) as a subgroup of G, then it occurs as many times as a subgroup of H. Thus, if we want to study the possible subgroups of a given group, we only have to pick out a suitable element of its equivalence class and to investigate the subgroups of this element.

More generally we write $\varepsilon G \prec \varepsilon G'$, if G can be isomorphically embedded in G', but the converse does not necessarily hold. This induces a partial order between the εG[2]).

We shall be interested in the equivalence class $\varepsilon S_\mathfrak{m}$ of $S_\mathfrak{m}$, the full permutation group of $\mathfrak{m}$ elements (hence $|S_\mathfrak{m}| = 2^\mathfrak{m}$). Recently DE BRUIJN [2] investigated the possible subgroups of $S_\mathfrak{m}$. It appears that quite a lot of groups of order $\leqq 2^\mathfrak{m}$, e.g. all such abelian groups and all such groups which are free or direct products of countable groups, are isomorphically contained in $S_\mathfrak{m}$. Therefore, the same holds for all groups of the equivalence class of $S_\mathfrak{m}$. Let us illustrate the concept of equivalence by a simple example.

Little or nothing is known about the structure of the automorphism groups $A(F_\mathfrak{m})$ and $A(V_\mathfrak{m})$ of the free group $F_\mathfrak{m}$ and the free abelian group $V_\mathfrak{m}$ ($F_\mathfrak{m}$ and $V_\mathfrak{m}$ both of order $\mathfrak{m}$). However, their systems of subgroups are the same and coincide with that of $S_\mathfrak{m}$ (and the results of DE BRUIJN can be applied):

$$\varepsilon S_\mathfrak{m} = \varepsilon A(S_\mathfrak{m}) = \varepsilon A(F_\mathfrak{m}) = \varepsilon A(V_\mathfrak{m}) .$$

Indeed, every permutation of a set of $\mathfrak{m}$ free generators of $F_\mathfrak{m}$ induces an automorphism of $F_\mathfrak{m}$, so $\varepsilon S_\mathfrak{m} \prec \varepsilon A(F_\mathfrak{m})$. Conversely, $F_\mathfrak{m}$ contains $\mathfrak{m}$ elements and every automorphism of $F_\mathfrak{m}$ permutes these elements, so $\varepsilon A(F_\mathfrak{m}) \prec \varepsilon S_\mathfrak{m}$. Hence $\varepsilon S_\mathfrak{m} = \varepsilon A(F_\mathfrak{m})$ and a corresponding proof holds for $V_\mathfrak{m}$.

It is easy to construct countable non-equivalent abelian groups. It is more difficult to prove that $\varepsilon S_{\aleph_1}$ and $\varepsilon S_{\aleph_1, \aleph_0}$ are incomparable (where $S_{\mathfrak{m},\mathfrak{n}} (\mathfrak{m} \geqq \mathfrak{n})$

---

[2]) From an announcement of JÓNSSON [15] it would follow — under the assumption of the generalized continuum hypothesis — that the set of εG's for the family $\{G\}$ with $|G| = \mathfrak{m}$ has a maximal element.

denotes the group of all permutations of $\mathfrak{m}$ elements which leave all elements fixed with the possible exception of a subset of cardinal $< \mathfrak{n}$). $S_{\aleph_1,\aleph_0}$ cannot be embedded into $S_{\aleph_0}$ by a theorem of DE BRUIJN [2], p. 567. $S_{\aleph_0}$ cannot be embedded into $S_{\aleph_1,\aleph_0}$, since in this last group every element has finite order.

As another application, we prove the relation

$$\varepsilon A(D_{\mathfrak{m}}) = \varepsilon S_{\mathfrak{m}}, \tag{4.1}$$

where the left-hand-side denotes the equivalence type of the (full) autohomeomorphism group of the generalized discontinuum $D_{\mathfrak{m}}$ (topological product of $\mathfrak{m}$ discrete spaces, each consisting of two points). The usefulness of (4.1) lies i.a. in the fact that it characterizes all possible groups $T(D_{\mathfrak{m}})$, simply by means of the fact that such a $T(D_{\mathfrak{m}})$ has to be a permutation group of $\mathfrak{m}$ objects.

It is clear that $\varepsilon S_{\mathfrak{m}} \prec \varepsilon A(D_{\mathfrak{m}})$, since every permutation of the $\mathfrak{m}$ pairs of points, used to define $D_{\mathfrak{m}}$, determines a topological map of $D_{\mathfrak{m}}$. Conversely, in order to prove $\varepsilon A(D_{\mathfrak{m}}) \prec \varepsilon S_{\mathfrak{m}}$ we proceed as follows.

$A(D_{\mathfrak{m}})$ is (STONE [22]) isomorphic to the automorphism group of the Boolean algebra corresponding to the (bicompact, 0-dimensional, Hausdorff) space $D_{\mathfrak{m}}$. This Boolean algebra contains exactly $\mathfrak{m}$ elements, and so its automorphism group can be considered as a subgroup of the full permutation group of these $\mathfrak{m}$ elements.

Relation (4.1) and theorem 5 below suggest that the $A(H)$ of an infinite, 0-dimensional, bicompact, Hausdorff space H might be fairly big in general. We know, however, that this is not the case, since there exist rigid H [14, 16, 20].

5. Groups of autohomeomorphisms of 0-dimensional spaces

First we mention some examples which prove to be fundamental in this area.

Observe that a homeomorphism maps both open and closed subsets onto both open and closed subsets.

Example I: Let R with $|R| = \varkappa$ be a 0-dimensional set of real numbers, rigid under continuous displacements, (let in theorem 1 M be the discontinuum of Cantor and R some F_γ) and R' a set of reals homeomorphic to R with $\overline{R}' \cap \overline{R} = \theta$. Put $N = R \cup R'$. We shall prove that $A(N)$ is isomorphic to the direct sum of continuously many groups, each of order two.

If $\varphi(N) = N$ and φ topological, we prove $\varphi^2 = e$. If $\varphi r = r_1 \in R$ for some r, we may conclude $r = r_1$, since R is rigid under continuous displacements. If $\varphi r = r' \in R'$ we may conclude $\varphi r' = r$ from the rigidity of R and R', as can easily be seen when considering φ and φ^2. Hence $\varphi^2 = e$. Such a group is necessarily abelian and has the structure as mentioned. It is clear that $\varepsilon A(N)$ is strictly smaller than $\varepsilon S_{\aleph_0}$.

Example II: Let $N = \bigcup\limits_{k=1}^{\infty} R_k$ and $N_m = \bigcup\limits_{k=1}^{m} R_k$ be sets of reals where the closures $\overline{R}_k$ are mutually disjoint and each of the R_k is homeomorphic to the set R of example I.

It is not difficult to prove that in N_m the orbit of a point under $A(N_m)$ consists of exactly m points, and that $A(N_m)$ is periodic. In fact, the order of every element is a divisor of $m!$. The group $A(N_\infty)$ is clearly not periodic, but contains elements of every order.

Let N denote some 0-dimensional separable metrizable space, abbreviated in this section by *0-space*. Regarding the normal subgroups of $A(N)$, it will be difficult to give a general analysis. In many important cases $A(N)$ is simple. R. D. Anderson e.g. proved[3]) that $A(N)$ is simple, if N either denotes the rationals or the discontinuum of Cantor. In a number of other cases there are only a few normal subgroups. e.g. if N is the set of integers (Schreier and Ulam [21]). $A(N)$ may be a direct product, e.g. if N is the union of two disjoint sets, the first homeomorphic to the rationals and the second homeomorphic to the irrationals, while both are understood to be open and closed in their union. By introducing mutually non-homeomorphic rigid spaces and taking suitable unions one can easily construct examples of 0-spaces for which the $A(N)$ decompose into unrestricted direct products with a denumerable number of factors; so we see already that the number of possibilities for the structure of $A(N)$ is abundant.

However, there are also properties which may be analysed in general for the $A(N)$, some of which we shall derive now. In particular we shall determine the structure of the center of $A(N)$.

(i) *If M is a 0-dimensional set of real numbers, dense in itself, then either $|A(M)| = 1$, or $A(M)$ contains continuously many elements each of order two.*

Proof. If M is topologically rigid (section 3), clearly, $|A(M)| = 1$. Thus, we may assume that M admits a non-trivial topological map φ: $\varphi M = M$. Suppose $\varphi(a) = b$, $a \neq b$, $a, b \in M$. Take disjoint neighbourhoods $U = U(a)$ and $V = V(b)$ in M, both open and closed (in M), such that $\varphi U = V$. If U is non-compact, U is, since it is dense in itself, the union of a countable number of mutually disjoint non-empty sets U_i, each U_i being open and closed in M. Let φ_i be the restriction of φ to U_i and $\varphi_i U_i = V_i$. For any set of positive integers n_i ($i = 1, 2, \ldots$), one can define an autohomeomorphism ψ of M by

$$\begin{aligned} \psi(p) &= \varphi_{n_i}(p) && (p \in U_{n_i};\ i = 1, 2, \ldots) \\ \psi(p) &= \varphi_{n_i}^{-1}(p) && (p \in V_{n_i};\ i = 1, 2, \ldots) \\ \psi(p) &= p && \text{elsewhere}. \end{aligned}$$

Since there are $\aleph$ different maps ψ of this kind, and each has order two, we proved our statement in the case that U is non-compact. However, if U is compact, it is homeomorphic to the discontinuum of Cantor, and then clearly, our contention holds.

The following corollaries are easy consequencies of (i).

(ii) *If K is a set of real numbers, dense in itself, then either $|A(K)| = 1$ or $|A(K)| = \aleph$.*

---

[3]) Quotation from a letter of 20. 2. '58.

(iii) *If M is a set of real numbers or a 0-dimensional separable metrizable space, then either $|A(M)| = n!$ for some positive integer n, or $|A(M)| = \aleph$.*

(iv) *If M is a 0-dimensional metric space, dense in itself, then either $|A(M)| = 1$ or $|A(M)| \geq \aleph$.*

Theorem 4: *If N is a 0-dimensional separable-metrizable space, then the center Z of $A(N)$ has one of the following structures.*

1°. *Z is trivial.*

2°. *$Z \simeq C_2$ (cyclic group of order two).*

3°. *Z is isomorphic to the restricted direct sum ΣC_2 of continuously many groups C_2.*

In order to prove this theorem and to obtain some understanding of the way in which $A(N)$ operates on N we introduce the following notions.

The *orbit* of $n \in N$ is, as usual, the set $\{\alpha n | \alpha \in A(N)\}$. $n \in N$ is *invariant*, if its orbit consists of $\{n\}$ itself. The set of all invariant points is called the *invariant set* F (F is closed in N). If a point $n \in N$ has a suitable open neighbourhood $U = U(n)$ (in N), such that the orbit of every point of U consists of exactly *two* points, such a pair of points (n, n') is called an involutory pair. The, apparently, open subset of N consisting of all involutory pairs is called the *involutory set* J. We restrict the notion of an *involution* φ to such elements $\varphi \in A(N)$ of order two which leave the points of $N \backslash J$ pointwise fixed. Observe that such a φ maps J necessarily onto itself. Observe furthermore, that not necessarily every homeomorphism of J onto itself may be extended to an involution (in suitably chosen examples).

If φ is an involution, one can prove (using some properties of 0-spaces and the Lindelöf covering theorem) that there exists a separation $J = J_1 \cup J_2$ (J_1 and J_2 disjoint and open in J), such that J_1 and J_2 are homeomorphic. Moreover, there are open and closed subsets of J_1 and J_2 respectively which are interchanged by φ, while φ leaves the other points of N pointwise fixed. The set J is uniquely determined by N. However, J_1 and J_2 are only topologically determined by N. The 0-space N being given, J_1 and J_2 may be chosen once for all, independent of φ. It follows from the definition of J that J_1 and J_2 are necessarily rigid.

Appendix to theorem 4: To each of the cases mentioned in theorem 4 corresponds the following operation of $A(N)$ on N.

1°. The invariant set F may be empty or not (and for any 0-space F one can construct examples of a 0-space N with $N \supset F$ such that F is its invariant set). J is empty and there are no involutions.

2°. F the same as in 1°. N contains exactly two isolated points, forming the involutory set J. There is one involution, generating Z.

3°. F is the same as in 1°. $J = J_1 \cup J_2$ where J_1 and J_2 are disjoint, homeomorphic, rigid, both open and closed in J, each consisting of more than one point. Z is generated by the involutions.

The following corollary is almost immediate.

Corollary: If N is some 0-dimensional separable metrizable space, the only possible abelian groups $A(N)$ are the following ones.

1°. $$A(N) \simeq \{e\}.$$

In this case N is necessarily rigid[4]).

2°. $$A(N) \simeq C_2.$$

N may consist of two points or of the union of two isolated points and a rigid space, dense in itself.

3°. $$A(N) \simeq \Sigma C_2.$$

N may consist of a closed invariant subset F (F may be any 0-dimensional separable metrizable space in suitable examples) and an involutory set $J = J_1 \cup J_2$. The sets J_1 and J_2 satisfy the properties as mentioned. In particular they are rigid.

Remark: Taking advantage of the results of SCHWEIGERT and FINE [5], one may obtain information on the center of $A(R)$, where R denotes an arbitrary set of reals.

Proof of theorem 4: Suppose $\zeta \in J$. Let $\zeta a \neq a$. We shall prove $a, \zeta a \in J$. Suppose $a \notin J$. Then, since also clearly $a, \zeta a \in F$, there is in every neighbourhood $U = U(a)$ a point c such that the orbit of c consists of at least three different points. We may assume $U \cap \zeta U = \theta$. Apart from c and ζc there is a point d different from both, which belongs to the orbit of c. Hence, there is a $\gamma \in A(N)$ being constant outside a small neighbourhood of c, in particular constant on ζc, such that $\gamma c = d$. Now apparently $\zeta c = \gamma \zeta c \neq \zeta \gamma c = \zeta d$, so $\zeta \notin Z$, which gives a contradiction. Hence $a, \zeta a \in J$. It follows that every element of Z unequal to the identity is necessarily an involution.

Conversely, if $\alpha \in A(N)$ and λ is an involution, one may write $\alpha = \mu . \nu$, where ν is an involution or the identity and μ leaves invariant all points which are non-invariant under λ. It follows that

$$\alpha \lambda = \mu \nu \lambda = \mu \lambda \nu = \lambda \mu \nu = \lambda \alpha .$$

So every involution λ belongs to Z.

Theorem 4 and the main part of the appendix now follow easily in view of the remarks made above concerning J and Z.

The remark that for any 0-space F one can construct a 0-space N with $N \supset F$ such that F is its invariant set is not important. It can be proved by exploiting the results on rigidity in section 3.

The following theorem gives information concerning the possible groups $T(N)$, using the results on equivalence of section 4. We give a completely elementary proof of this theorem not using Boolean algebras as in the proof of (4.1).

---

[4]) The rigidity involves $|N| = 1$ or $\aleph$, if we assume the continuum hypothesis. Observe furthermore, that a rigid 0-dimensional space may contain one isolated point.

Theorem 5. *If N either denotes a countably infinite metrizable space or an infinite locally compact 0-dimensional separable, metrizable space, then*

$$\varepsilon A(N) = \varepsilon S_{\aleph_0}. \tag{5.1}$$

Proof: Suppose N is countably infinite and metrizable. Every homeomorphism of N permutes this countable set. It easily follows that

$$\varepsilon A(N) \prec \varepsilon S_{\aleph_0}. \tag{5.2}$$

Conversely, if N contains an infinite number of isolated points, we may determine a sequence of such points which either converges in N or has no limit point at all. Then every permutation of the points of this sequence can be extended to an autohomeomorphism of N, so

$$\varepsilon S_{\aleph_0} \prec \varepsilon A(N). \tag{5.3}$$

However, the same relation holds if N contains only a finite number of isolated points, since in this case we may determine a sequence of subsets of N each of which is open and closed in N and dense in itself (hence homeomorphic to the space of rationals), while the union of the elements of the sequence is also open and closed in N. So we can permute the elements of this sequence, generating the required group of autohomeomorphisms.

If N is infinite, locally compact, 0-dimensional and separable metrizable, the relation (5.3) follows in about the same way. Indeed, if N contains only a finite number of isolated points, N contains an open and closed subset D homeomorphic to the discontinuum of Cantor and since D is the topological product of a countable number of pairs of points, (5.3) follows.

The only non-trivial case is to prove (5.2). We compactify by one point to a 0-dimensional compactum N^*. $A(N)$ is clearly a subgroup of $A(N^*)$. It can easily be seen that the system of both open and closed subsets of N^* is countable. Every element of $A(N^*)$ permutes the elements of this system; to different elements correspond different permutations of the system and to the product of two elements corresponds the product of the corresponding permutations. So (5.2) follows, if we interpret $S_{\aleph_0}$ as the full permutation group of the system of both open and closed subsets of N^*.

The 0-space N in example I and N_m in example II are 0-spaces with $|N| = |N_m| = \varkappa$ and $|A(N)| = |A(N_m)| = \varkappa$ such that nevertheless the equivalence types of $A(N)$ and $A(N_m)$ are strictly smaller than $\varepsilon S_{\aleph_0}$. On the other hand $\varepsilon A(N_\infty) = \varepsilon S_{\aleph_0}$, where N_∞ is the 0-space of example II. Indeed, let ψ: $\psi R_j = R_i$ be the natural topological map of R_j onto R_i. If φ is topological and $\varphi r_i = r_j$, $r_i \in R_i$, $r_j \in R_j$, then $\psi\varphi$ must map r_i onto itself, since R_i is rigid under continuous displacements. So, for every φ a point $r_i \in R_i$ is mapped on $\psi^{-1} r_i = r_j$ for some j. This means that, if D_1 is some countable set, dence in R_1, and D_j the corresponding subset of R_j under the natural map of R_1 on R_j, any φ must permute the points of the countable set $\bigcup_{k=1}^{\infty} D_k$. Hence $\varepsilon A(N) < \varepsilon S_{\aleph_0}$. Conversely, $\varepsilon S_{\aleph_0} < \varepsilon A(N)$, since every permutation of the R_k induces in the natural way a topological map of N onto itself. Hence $\varepsilon A(N) = \varepsilon S_{\aleph_0}$.

More generally, if one takes the union K of $\mathfrak{m}$ disjoint copies of R, we may obtain a 0-dimensional metrizable space K, dense in itself, with $|K| = \mathfrak{m} \cdot \aleph$, such that $\varepsilon A(K) = \varepsilon S_{\mathfrak{m}}$.

Generally speaking, we see that the $A(N)$ of sets N produce a smaller family (though we have shown them to possess a certain diversity) than the family of the $A(P)$ for all 1-dimensional (or n-dimensional) separable connected metric spaces P. Indeed, every countable group is isomorphic to an $A(P)$ for some Peano-curve P (cf. [13]). Conversely, for every N there exists such a P with $A(N) \simeq A(P)$. To prove this, take e.g. the topological product $N \times H$ of N and a suitable topologically rigid, connected, locally connected plane set H. P is obtained by identifying the subset $N \times h$ ($h \in H$ fixed) of $N \times H$ to one and the same point.

There remain some open problems regarding G_δ-sets.

Problem: Does there exist a topologically rigid, infinite G_δ-subset of the real line?

Problem: Give essential information on the equivalence type of the space of irrational numbers L.

The author does not know e.g. whether εL is actually greater than $\varepsilon S_{\aleph_0}$.

6. Graph-representations of groups

We recall the notion of a *Cayley-graph*[5]). This is a representation of a group G by means of a coloured, directed graph G^*. Let $\{s_\alpha\}$ be a system of generators of G. Each element of G is represented by one vertex of G^*. It is convenient to use the same symbols $a, b, \ldots, s, \ldots$ for the elements of G and the corresponding vertices in the graph. With each generator s_α we associate a certain set of directed edges. Two vertices, a and b, are joined by an s_α-edge, directed from a to b, whenever

$$b = s_\alpha a \,. \tag{6.1}$$

The edges join the vertices to their "neighbours"; the neighbours of the unit element e represent the generators s_α and their inverses. G^* is clearly connected. With each $a \in G$ we associate the right translation π_a:

$$\pi_a : x \to xa \quad (x \in G) \,.$$

When applying products of mappings from the left to the right

$$(x)\pi_a\pi_b = (x\pi_a)\pi_b \,,$$

we see that G is clearly isomorphic to its right regular representation G_r. It can easily be seen that G_r is isomorphic to the group of all permutations of the vertices of G^* which are colour and orientation preserving. Hence G *is iso-*

[5]) Cayley defined his graph for the case of finite groups (1878), but exactly the same definition applies to the infinite case. Dehn, who rediscovered this representation and gave interesting applications for the case of groups with a finite number of generators, shows an outspoken tendency [4], **69**, p. 144 to minimize Cayley's contribution. See Coxeter and Moser [3] for more details on Cayley-graphs.

morphic to the automorphism group of the coloured, directed graph G^*. Observe that every automorphism, unequal to the identity, is fixed-point-free.

Now we shall extend these results in the following way: we define for any given group an (uncoloured) graph $\mathfrak{G}$ such that the automorphism group $A(\mathfrak{G})$ of $\mathfrak{G}$ is isomorphic to G. This has been done for the case of *finite* groups by FRUCHT [6] (see also [7]).

First we construct a *rigid graph*, i.e. a graph for which the automorphism group is the identity. The construction will be carried out in a countable number of steps and uses induction in n.

First step: Take a cardinal $\mathbf{m}$. Take a segment (edge) $[pq]$ and draw from q (one of its vertices) $\mathbf{m}$ new segments $[q, q_{\alpha_1}]$ in pairs disjoint, with the exception of the endpoint q which they have in common. Suppose after the n-th step a graph has been constructed of the following nature: $[pq]$, $[qq_{\alpha_1}]$, $[q_{\alpha_1}q_{\alpha_1\alpha_2}], \ldots,$ $[q_{\alpha_1\alpha_2\cdots\alpha_{n-1}}q_{\alpha_1\alpha_2\cdots\alpha_{n-1}\alpha_n}]$, such that for $i < n$ the vertex $q_{\alpha_1\alpha_2\cdots\alpha_i}$ defined as vertex of $[q_{\alpha_1\alpha_2\cdots\alpha_{i-1}}q_{\alpha_1\alpha_2\cdots\alpha_{i-1}\alpha_i}]$ is vertex of $\mathbf{m}_{\alpha_1\alpha_2\cdots\alpha_i\alpha_{i+1}}$ different edges $[q_{\alpha_1\alpha_2\cdots\alpha_i}q_{\alpha_1\alpha_2\cdots\alpha_i\alpha_{i+1}}]$, while two cardinals $\mathbf{m}_{\alpha_1\alpha_2\cdots\alpha_i}$ $(i \leq n)$ which differ in at least one of their subindices, will be different. Moreover, the vertices $q_{\alpha_1\alpha_2\cdots\alpha_n}$ will be vertices of order 1. We now define the graph of the $(n+1)$-th step. Attach to each such q of order 1 a cardinal $\mathbf{m}_{\alpha_1\alpha_2\cdots\alpha_n\alpha_{n+1}}$. All these will be different from each other and from all the $\mathbf{m}$-cardinals previously defined. Now draw from each $q_{\alpha_1\alpha_2\cdots\alpha_n}$ $\mathbf{m}_{\alpha_1\alpha_2\cdots\alpha_n\alpha_{n+1}}$ new segments, disjoint with the possible exception of $q_{\alpha_1\alpha_2\cdots\alpha_n}$. This new graph has the same properties as mentioned for the graph of the n-th step (n being replaced by $n+1$) — as can easily be seen. The construction is completed after a countable number of steps. The rigidity of the resulting graph can easily be deduced and depends on the fact that for each segment (edge) the pair of orders of its endpoints (vertices) is unique in the graph. Moreover, it is not difficult to see that for every cardinal $\mathbf{k}$, we can construct $\mathbf{k}$ rigid graphs, such that no pair of them is isomorphic.

Now let G^* be a coloured directed graph of the group G in which all elements unequal to e are taken as generators. We assume $\mathbf{k} = |G| > 3$ (the case of groups of order ≤ 3 can be treated separately in a simple way), so the graph G^* has an order ≥ 3 in any of its vertices.

Consider all directed edges of a fixed colour s in G^*. Construct a rigid graph R as described above, such that all the $\mathbf{m}$-cardinals occurring in the definition of R are greater than $\mathbf{k}$. The connected graph R' is defined as follows: edges $[ar]$, $[rp]$, $[pb]$, and furthermore R, the vertex of order 1 of R identified with the vertex p in this definition. Suppose that a directed edge with colour s leads from vertex a to vertex b in G^*. Replace this directed s in G^* by R', the a's and b's being retained. We do this for every such s, replacing them by mutually isomorphic R'. Furthermore, we repeat this construction for the other coloured edges of G^*, but we take care that the rigid $R = R_t$ for the colour t is non-isomorphic to the rigid $R = R_s$ for the colour s, distinct from t.

In this way G^* is transformed into a graph $\mathfrak{G}$, which satisfies our requirements, as can be seen by elementary deductions. Thus we have proved

Theorem 6. *Every group G is isomorphic to the automorphism group of some graph $\mathfrak{G}$. We may assume that this automorphism group operates fixed-point-free on $\mathfrak{G}$.*

Comments: König [17], p. 5 raised the question, which groups can be represented as automorphism group of graphs. As remarked before, Frucht [6] gave a solution for finite groups using also Cayley-graphs. The first general solution of König's problem can be deduced from the results given by Birkhoff [1]. Indeed, one of his results states that every group is isomorphic to the automorphism group of some distributive lattice.

If the order $|G|$ of G is given, the author does not know the smallest possible order $|\mathfrak{G}|$. However, if $|G| = \mathfrak{m}$, we may assume $|\mathfrak{G}| \leq 2^{\mathfrak{m}}$ (cf. Birkhoff [1]). See also Frucht [7] for the case of finite $|G|$. If $|G|$ is finite $|\mathfrak{G}|$ can be made finite (Frucht); if $|G| = \aleph_0$, their $\mathfrak{G}$ can be constructed such that $|\mathfrak{G}| = \aleph_0$ (and such that the order in a vertex of $\mathfrak{G}$ is at most three; finite case, Frucht [8]). *If $|G| \leq \aleph$ there is a $\mathfrak{G}$ with $|\mathfrak{G}| \leq \aleph$.* Let us indicate a proof of this contention in case $|G| = \aleph$. Take all elements of G unequal to e as generators and construct the corresponding Cayley-graph G^*. From G^* we construct $\mathfrak{G}$ in the same way as in the proof of the theorem above, apart from the fact that we use other rigid graphs R, defined as follows. To every characteristic function f defined on the natural numbers n with $f(0) = f(1) = 0$ we attach the following graph R: join the vertices $1, 2, \ldots, n, \ldots$ consecutively by edges $[n, n+1]$; furthermore, if $f(n) = 1$ for a certain n, draw a new edge from n as vertex, the other vertex of this edge remaining of order 1; if $f(n) = 0$, no alteration is made. In this way we obtain continuously many mutually non-isomorphic rigid graphs; for different colours we use again different R's. Hence, if $|G| = \aleph$, then $|\mathfrak{G}| = \aleph$.

Corollary: *Every group G is isomorphic to the automorphism group of some directed graph $\mathfrak{H}$.*

Proof: Take an R', as defined in the proof of theorem 5, and give every edge of R' some orientation, arbitrarily chosen. This orientation induces an orientation in all R', isomorphic to R'. Carry out this procedure over the entire $\mathfrak{G}$. The orientated $\mathfrak{G}$ obtained is the $\mathfrak{H}$, satisfying the requirements of the corollary, as can be proved.

7. Autohomeomorphism groups

Given some group G and the corresponding graph $\mathfrak{H}$ of the corollary in the preceding section, we shall replace the edges of $\mathfrak{H}$ by mutually homeomorphic, topologically rigid spaces H and introduce a metric in the resulting set, such that a space M will be obtained satisfying the following theorem.

Theorem 7. *Every group G is isomorphic to the autohomeomorphism group $A(M)$ of some 1-dimensional, connected, locally connected, complete metric space M.*

Moreover, the group of isometries of our M coincides with $A(M)$ and operates in a fixed-point-free way on M.

If $|G| \leq \aleph$, there is such an M with $|M| = \aleph$.

Proof. If we do not claim M to be complete, we can use for H a topologically rigid space, defined in section 3 (applying in the proof lemma 4 and corollary 2). To render M complete, we use the rigid curve P_n, defined at the end of section 3. If G is finite, of order m, we take care that $n > m$ and in any case $n > 3$. Let us denote such a P_n by H. We observe that H satisfies the following property (p): every point $h \in H$ has arbitrarily small neighbourhoods U such that $U \setminus \{h\}$ is either connected or contains more than n components.

In H we take two points ("vertices") h^0 and h^1, diametrically opposite on the outer circle, which will have a diameter one.

Now we replace each directed edge $[q^0 q^1]$ of $\mathfrak{H}$ (corollary, section 6) by such an H, h^i replacing q^i $(i = 0, 1)$. All H's are isomorphic to each other and disjoint with the possible exception of their "vertices". Into the union of all H's

$$M = \bigcup_\alpha H_\alpha$$

we introduce a metric ϱ. This metric will by definition coincide with the metric already given in each $H = H_\alpha$.

If $a \in H_\alpha$, $b \in H_\beta$ $(\alpha \neq \beta)$, then there is a finite "chain" of points, connecting a and b,

$$a,\ h^i_\alpha = h^{j_1}_{\gamma_1},\ h^{k_1}_{\gamma_1} = h^{j_2}_{\gamma_2},\ h^{k_2}_{\gamma_2} = h^{j_3}_{\gamma_3}, \ldots, h^{k_n}_{\gamma_n} = h^s_\beta,\ b \tag{1}$$

with $i, s = 0$ or 1; $j_t + k_t = 1$ for all $1 \leq t \leq n$; n is a natural number. We define the *length* λ of chain (1) by

$$\lambda = \varrho(a, h^i_\alpha) + \varrho(h^s_\beta, b) + n - 1\,.$$

The *minimal length* of all chains connecting a and b is defined as their distance $\varrho(a, b)$ in M. This function ϱ, now generally defined over $M \times M$, is indeed a distance function and it can easily be shown that the resulting metric space M is complete, connected and locally connected.

To prove the 1-dimensionality of M we must show that for an arbitrary closed set $C \subseteq M$ and an open set U containing C, there exists an open set $V \supset C$ such that $C \subset \overline{V} \subset U$ and $\dim(\overline{V} \setminus V) = 0$.

If $c \in C$ is not a vertex of an H, there is a neighbourhood $N(c)$ of c contained in H with $\overline{N(c)} \subset U$ and $\dim(\overline{N} \setminus N) = 0$, since H is a curve. If $h \in C$ is a vertex (of a number of H's) there is a neighbourhood $N = N(h)$ of diameter $< 1/2$ with $\overline{N(h)} \subset U$, and $\dim(\overline{N} \setminus N) = 0$, since the boundary of $N(h)$ can be taken as the sum of the mutually disjoint boundaries of $N \cap H$ for each H having h as a vertex.

Since H is compact, $C \cap H$ is covered by at most two sets $\overline{N(h)} \cap H$ together with a finite number of $\overline{N(c)}$. The boundary of the union of these sets is 0-dimensional. Carrying out this procedure for every H which intersects C, and taking the union of all the sets (contained in such a finite covering) under consideration, we obtain the required $\overline{V}$. The boundary of $\overline{V}$ is 0-dimensional, since it is the union of a number of 0-dimensional sets, each of which is open and closed in this union.

Next we prove $A(M) \simeq G$. If $\varphi M = M$ is an autohomeomorphism, we show that for any α there exists a β such that

$$\varphi H_\alpha = H_\beta \,. \tag{2}$$

Indeed, φH_α cannot contain points of two different H's. Otherwise it would contain a vertex h of some H. However, h is locally a cutpoint in M in such a way that if $N(h)$ is a suitable small neighbourhood of h in M, the set $N(h)\setminus\{h\}$ contains either an infinite number of components or less than $max(m, 3)$ components. This contradicts property (p) valid for H, since φ is topological. The same reasoning applied to φ^{-1} shows that φH_α cannot be properly contained in an H_β. Hence (2) holds.

From (2) it follows that φ permutes the H's among themselves, but then we may conclude from Theorem 6 and its corollary and the rigidity of H that $G \underset{T}{\simeq} A(\mathfrak{H}) \simeq A(M)$. Moreover, it is easy to see that such a φ is an isometry of M. So every autohomeomorphism of M is an isometry and the corresponding groups coincide. Finally, suppose $|G| \leq \varkappa$. Then there is a $\mathfrak{G}$ and also an $\mathfrak{H}$ with $|\mathfrak{G}|, |\mathfrak{H}| \leq \varkappa$ (see preceding section). Hence $|M| = \varkappa^2 = \varkappa$ for the corresponding M.

Remark. Replacing H by a suitable rigid space of some suitable dimension, it can be proved that the metric space M in Theorem 7 may have any positive dimension, finite or infinite.

Applying the Čech-Stone bicompactification theory on the preceding theorem the following result is easily obtained.

Theorem 8. *Every group is isomorphic to the autohomeomorphism group of some connected bicompact Hausdorff-space.*

Proof. Given the group G, construct the metric space M, such that $A(M) \simeq G$. If $\tilde{M}$ is the Čech-Stone bicompactification of M the following holds true: every autohomeomorphism $\varphi M = M$ can be uniquely extended to an autohomeomorphism $\tilde{\varphi} \tilde{M} = \tilde{M}$. Conversely, if $\psi \tilde{M} = \tilde{M}$ and ψ is topological, then necessarily (ČECH) $\psi M = M$, since the metric space M satisfies the first axiom of countability. From this follows easily

$$A(\tilde{M}) \simeq A(M)\,; \quad \text{so} \quad A(\tilde{M}) \simeq G\,,$$

which we had to prove.

8.1. Automorphism groups of Boolean rings

If B is a Boolean ring it is known [22] that the automorphism group $A(B)$ is isomorphic to the automorphism group of the corresponding Boolean algebra and to the autohomeomorphism group $A(N)$ of the corresponding Boolean space N.

We also recall the statements, used at the end of the previous section: If M is v_0-dimensional, metrizable, then autohomeomorphisms of M induce autohomeomorphisms of βM, the maximal 0-dimensional compactification of M, and conversely, in such a way that $A(M) \simeq A(\beta M)$.

Now we can apply some of the results of sections 3, 4 and 5 to obtain some information on the groups $A(B)$.

Indeed, if M is a (0-dimensional) rigid subset M of the real line, dense in itself, the Boolean space $\beta M = N$ is also topologically rigid. We can even choose M in such a way that it does not admit non-trivial continuous maps onto or topological maps into itself and then the same holds for the Boolean space $\beta M = N$. It is not difficult to translate these properties into the terminology of Boolean algebras. In particular, it shows the *existence of rigid Boolean rings* B (of continuous potency, apparently), i.e. rings B for which $A(B)$ equals the identity. This answers again in the negative a question raised by BIRKHOFF. Previous solutions have been given by JÓNSSON [14], RIEGER [20] and KATĚTOV [16]. However, we can derive stronger results. Translation of theorem 2 into the language of Boolean algebras leads immediately to the following result.

Corollary: (to theorem 2): *There exists a family $\{B_\gamma\}$ of $2^\varkappa$ Boolean rings B_γ, each of potency $\varkappa$, such that no B_γ can be mapped homomorphically onto any other B_γ or non-trivially onto itself.*

RIEGER [20] found already an example of a Boolean algebra which admitted no proper homeomorphism onto itself, but its cardinal is "unsatisfactorily high", as RIEGER remarked.

If M is the union of two homeomorphic rigid disjoint subsets of the line, dense in itself, each closed in their union, then the $A(B)$ of B corresponding to βM is an infinite group in which every element $(\neq e)$ has order two.

These examples are typical for the situation in general, as is shown by the following theorem.

Theorem 9. *If B is a countably infinite Boolean ring, then its automorphism group $A(B)$ is equivalent to $S_{\aleph_0}$. If, however, B is an atomless Boolean ring of at least continuum order then $A(B)$ is either the identity-group or $A(B)$ is infinite, actually contains an infinite number of elements each of order two.*

This theorem contradicts some results of RIEGER [20] p. 214, not affecting the main results of his paper, in which he claimed the construction of atomless Boolean algebras B for which $A(B)$ is isomorphic to the symmetric group of some finite number of elements.

It also serves as a counter-part to Birkhoff's theorem [1], as already mentioned in section 6, in which he shows that *every* group is isomorphic to the $A(L)$ of some *distributive* lattice L. Indeed, theorem 9 shows in particular that many groups cannot be represented as groups $A(B)$, where B is some Boolean algebra, i.e. a *complemented distributive* lattice. It would be useful to have more information on the $A(B)$; e.g., for which cardinals do there exist rigid B? Can there be said more on the cardinal of $A(B)$ for a given B? Can the center of $A(B)$ be determined in a way similar to theorem 4?

Proof of theorem 9: If B is countably infinite, the corresponding Boolean space N has a countable number of both open and closed subsets. So N has a countable base. Since N is also a compact Hausdorff space, N must be a

0-dimensional, compact, metrizable space. Applying theorem 5, we see that $A(N)$ and hence $A(B)$ is equivalent to $S_{\aleph_0}$.

The second part of the theorem also easily follows in a way similar to the proof of section 5 (i). Indeed, if N (corresponding to B) is not rigid, there is an autohomeomorphism φ of N which maps an open and closed subset U of N onto a disjoint φU. Now we can decompose U into an arbitrary finite number of both open and closed, disjoint subsets $V_1, V_2, \ldots, V_n$. Interchanging between some of the V_i and φV_i and leaving constant the rest of N, leads to autohomeomorphisms of N, each of order two. Since there is an apparently infinite number of such maps, all different, the theorem has been proved.

8.2. Automorphism groups of rings

The results of section 7 admit a purely algebraic translation by applying a fundamental result of GELFOND and KOLMOGOROV [9] (cf. also M. H. STONE [22]) on rings of continuous real functions defined on completely regular spaces.

Theorem 10. *Every group G is isomorphic to the automorphism group of some ring R, actually of infinitely many commutative rings with unit-element.*

If $|G| \leq \aleph_0$ or $|G| \leq \varkappa$, then there are such R with $|R| = \varkappa$ or $|R| = 2^\varkappa$ respectively.

Observe furthermore: If $|G| = \mathfrak{m}$, R can be made such that $|R| \leq 2^{2^{\mathfrak{m}}}$ (see the comments in section 6).

If one asks for those groups which are automorphism groups $A(G)$ of groups G instead of rings R, there is no such simple answer. Indeed, it is well known that the cyclic groups of odd order are no $A(G)$; also "most" of the cyclic groups of order $8n$ are no $A(G)$; the dihedral group D of order 8 is no automorphism group of a group, with the exception of itself: $D \simeq A(D)$. See DE VRIES and DE MIRANDA [24] for these and a number of other results concerning the automorphism groups of groups.

Some of their results can easily be generalized for topological groups as well. In particular a given group is in general not isomorphic to the automorphism group of a non-abelian topological group (e.g. a non-abelian topological group, unequal to the symmetric group S_3, has at least 8 automorphisms. There exist countable, non-abelian, non-discrete groups with exactly 8 automorphisms). Theorem 10 asserts in particular that every group is isomorphic to the automorphism group of some topological commutative ring with a discrete space-structure. Let us mention without proof that one can also introduce a non-discrete topology in the particular ring to be defined, without changing its automorphism group.

It is unknown whether there exist abelian groups G the order of which is an arbitrary infinite cardinal, such that $A(G) \simeq C_2$, where C_2 is the cyclic group of order two; in particular it is unknown whether there exist indecomposable abelian G of very large cardinals. However, there exist commutative rings R of large potency such that $A(R) \simeq C_2$, or more generally $A(R) \simeq H$,

for any group H. This follows from our proof of Theorem 10. Indeed, if we use graphs of high potency for the rigid graphs defined in section 6, this will result in a ring of high order.

Regarding those groups which can be automorphism groups of divisionrings or fields, the author did not investigate them and must confess his ignorance. A superficial investigation has not even revealed a solution of the following very special *problem*: Is the infinite cyclic group the automorphism group of some field?

The *proof of the main part of Theorem 9*: clearly follows from the lemma below. Indeed, we can either apply the first part of the lemma on Theorem 7 or the second part on Theorem 8.

Lemma 5: If S is a completely regular space satisfying the first axiom of countability, and $C(S)$ is the ring of all continuous real-valued functions defined on S, then $A(S) \simeq A(C(S))$.

If B is a bicompact Hausdorff-space and $C(B)$ its ring of c. r. functions, then $A(B) \simeq A(C(B))$.

The *first part of this lemma* is contained in a paper of PURSELL [19] (theorem 4.6, p. 968. No proof is given).

The *second part of the lemma* will be contained in a forthcoming book by L. GILLMAN and M. JERISON [10]. Let us for the sake of completeness indicate how it can easily be deduced from the fundamental paper [9]. By using

$$C(B)/J(b) \simeq K ,$$

where K is the field of reals and $J(b)$ a maximal ideal in $C(B)$ (consisting of all functions vanishing at a certain $b \in B$), and by using the *rigidity* of K (as a field), it follows that every automorphism of $C(B)$, unequal to the identity, determines a permutation, unequal to the identity, of the set of maximal ideals $\{J(b)\}$.

Every autohomeomorphism φ of B clearly induces an automorphism φ' of $C(B)$. Conversely, every automorphism φ' of $C(B)$ determines a permutation of $\{J(b)\}$ and induces an autohomeomorphism of B. Using the remark made above, one can conclude that this autohomeomorphism must be φ, from which the second part of the lemma follows.

The *addendum on cardinals in Theorem* 10 follows in case $|G| \leq \aleph_0$ from the result in [13], i.e. there is a compact metric, hence separable, M such that $G \simeq A(M)$. The second part is obvious in view of the addendum of Theorem 7.

References

[1] BIRKHOFF, G.: Sobre los grupos de automorfismos. Rev. Union mat. Argent. **11**, 155—157 (1945). — [2] BRUIJN, N. G. DE: Embedding theorems for infinite groups. Indag. Math. **19**, 560—569 (1957). — [3] COXETER, H. S. M., and W. O. J. MOSER: Generators and relations for discrete groups. Berlin 1957. — [4] DEHN, M.: Math. Ann. **69**, 137—168 (1910); **71**, 116—144 (1912). — [5] FINE, N. J., and G. E. SCHWEIGERT: On the group of homeomorphisms of an arc. Ann. Math. **62**, 237—253 (1955). — [6] FRUCHT, R.: Herstellung von Graphen mit vorgegebener abstrakter Gruppe. Comp. Math. **6**, 239—250 (1938). — [7] FRUCHT, R.: Sobre la construccion de sistemas parcialmente ordenados con grupo

de automorfismos dedo. Rev. Union mat. Argent. 13, 12—18 (1948). — [8] Frucht, R.: Graphs of degree three with a given abstract group. Canad. J. Math. 1, 365—378 (1949). — [9] Gelfond, I., and A. Kolmogorov: On rings of continuous functions on topological spaces. Doklady Acad. Sci. USSR 22, 11—15 (1939). — [10] Gillman, L., and M. Jerison: Rings of continuous functions. New York: Van Nostrand. To appear. — [11] Groot, J. de: Continuous mappings of a certain family. Fund. Math. 42, 203—206 (1955). — [12] Groot, J. de: Automorphism groups of rings. Abstract of short communication Int. Congr. of Math., Edinburgh 1958, 18. — [13] Groot, J. de, and R. J. Wille: Rigid continua and topological group-pictures. Archiv der Math. 9, 441—446 (1958). — [14] Jónsson, B.: A Boolean algebra without proper automorphisms. Proc. Amer. math. Soc. 2,766—770 (1951).— [15] Jónsson, B.: Universal relational systems (abstract). Bull. Amer. math. Soc. 64, 403 (1956).— [16] Katětov, M.: Remarks on Boolean algebras. Coll. Math. 2, 229—235 (1951).— [17] König, D.: Theorie der endlichen und unendlichen Graphen. Leipzig 1936. — [18] Kuratowski, C.: Topologie. I. (édition troisième), Warszawa 1952. — [19] Pursell, L. E.: An algebraic characterization of fixed ideals in certain function rings. Pacific J. Math. 6, 963—969 (1956). — [20] Rieger, L.: Some remarks on automorphisms in Boolean algebras. Fund. Math. 38, 209—216 (1951). — [21] Schreier, J., and S. Ulam: Über die Permutationsgruppe der natürlichen Zahlenfolge. Studia math. 4, 134—141 (1933). — [22] Stone, M. H.: Applications of the theory of Boolean rings to general topology. Trans. Amer. math. Soc. 41, 375—481 (1937). — [23] Vries, H. de: Topological groups with a small number of automorphisms. To be published — [24] Vries, H. de, and A. B. de Miranda: Groups with a small number of automorphisms. Math. Z. 68, 450—464 (1957).

(Eingegangen am 13. Februar 1959)

GOTTSCHLING, E.
Math. Annalen 138, 103—124 (1959)

Explizite Bestimmung der Randflächen des Fundamentalbereiches der Modulgruppe zweiten Grades*

Von
ERHARD GOTTSCHLING in Göttingen

Die in der Einleitung zunächst zusammengestellten Ergebnisse sind in den beiden Arbeiten [1] und [2] von C. L. SIEGEL zu finden.

Für jedes natürliche n werde der Bereich aller n-reihigen komplexen symmetrischen Matrizen $Z = X + iY$, deren Imaginärteil Y positiv definit ist, mit $\mathfrak{H}$ bezeichnet. Die allgemeinste umkehrbar eindeutige Abbildung von $\mathfrak{H}$ auf sich, die in jedem der $\frac{1}{2}n(n+1)$ unabhängigen Elemente z_{kl} von Z $(k, l = 1,\ldots,n)$ analytisch ist, hat die Gestalt

$$W = (AZ + B)(CZ + D)^{-1}. \tag{1}$$

Dabei sind A, B, C, D reelle n-reihige Matrizen, die sich in der Form

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \tag{2}$$

zu einer $2n$-reihigen symplektischen Matrix zusammenfassen lassen. Ist E die n-reihige Einheitsmatrix und $J = \begin{pmatrix} 0 & E \\ -E & 0 \end{pmatrix}$, so bedeutet die Symplektizität von M die Bedingung

$$MJM' = J. \tag{3}$$

Zwei symplektische Matrizen M und M_1 liefern genau dann dieselbe Abbildung von $\mathfrak{H}$ auf sich, wenn $M_1 = \pm M$ ist. Die homogene Modulgruppe n-ten Grades besteht aus allen $2n$-reihigen symplektischen Matrizen mit ganzzahligen Elementen. Die inhomogene Modulgruppe Γ ist dann die Faktorgruppe der homogenen Modulgruppe nach der Untergruppe, die aus $\pm E_{2n}$ besteht, wobei E_{2n} die $2n$-reihige Einheitsmatrix ist. Wird Γ aufgrund von (1) als Gruppe von Abbildungen von $\mathfrak{H}$ auf sich gedeutet, so ist sie in $\mathfrak{H}$ diskontinuierlich. Es erhebt sich die Frage nach einem Fundamentalbereich von Γ in $\mathfrak{H}$. Diese Frage wurde zuerst von C. L. SIEGEL behandelt. Um sein Ergebnis zu formulieren, wird der Begriff des teilerfremden symmetrischen Matrizenpaares benötigt: Ein Paar von zwei n-reihigen Matrizen C und D mit ganzzahligen Elementen heißt symmetrisch, wenn $CD' = DC'$ gilt. Es heißt teilerfremd, wenn aus der Voraussetzung, die Elemente von GC und GD seien ganzzahlig, stets folgt, daß die n-reihige Matrix G selbst ganzzahlige Elemente besitzt. Die teilerfremden symmetrischen Matrizenpaare stehen in enger Beziehung zu

*) Diese Arbeit wurde von der Mathematisch-Naturwissenschaftlichen Fakultät der Universität Göttingen als Dissertation angenommen.

den Modulmatrizen: Aus (3) folgt nämlich, daß z. B. die zweite Matrizenzeile (CD) einer jeden Modulmatrix (2) symmetrisch und teilerfremd ist. Umgekehrt kann man auch jedes teilerfremde symmetrische Paar C, D zu einer Modulmatrix (2) ergänzen.

Im folgenden wird die Determinante einer komplexen Matrix W mit $\det W$ bezeichnet, der Absolutbetrag einer komplexen Zahl z mit $|z|$. Für den Absolutbetrag $|\det W|$ der Determinante von W soll kurz $\operatorname{abs} W$ geschrieben werden.

Der Bereich $\mathfrak{F}$ bestehe nun aus allen Punkten $Z = X + iY$ von $\mathfrak{H}$, die den folgenden Bedingungen genügen:

(4) $\frac{1}{2} \geqq x_{kl} \geqq -\frac{1}{2}$, *wobei* $X = (x_{kl})$ $(k, l = 1, \ldots, n)$;

(5) Y *nach* Minkowski *reduziert*;

(6) $\operatorname{abs}(CZ + D) \geqq 1$ C, D *teilerfremd und symmetrisch*.

Dann ist $\mathfrak{F}$ ein Fundamentalbereich von Γ in $\mathfrak{H}$. Um zu entscheiden, wann zwei der Ungleichungen (6) verschiedene Bedingungen für Z darstellen, werden die teilerfremden symmetrischen Paare noch in Klassen von zueinander assoziierten Paaren eingeteilt: Zwei Paare C, D und C_1, D_1 heißen assoziiert, wenn es eine n-reihige unimodulare Matrix U mit der Eigenschaft

$$(C_1 D_1) = U(CD)$$

gibt. In [2] wird gezeigt, daß die Beziehung

$$\operatorname{abs}(C_1 Z + D_1) = \operatorname{abs}(CZ + D)$$

genau dann identisch in Z gilt, wenn C, D und C_1, D_1 assoziiert sind. Also liefert (6) genau so viele verschiedene Ungleichungen, wie es Klassen assoziierter Paare C, D gibt. Diejenigen Ungleichungen (6), für die C, D zu $0, E$ assoziiert ist, stellen allerdings keine Bedingung für Z dar und sollen daher fortgelassen werden. Da es unendlich viele Klassen assoziierter Paare C, D gibt, und da die Zahl der Minkowskischen Reduktionsbedingungen ebenfalls unendlich groß ist, ist $\mathfrak{F}$ zunächst durch unendlich viele Ungleichungen definiert. Nun beweist aber Minkowski in der Arbeit [3] oder auch Siegel in der Arbeit [4], daß sich sämtliche Reduktionsbedingungen schon aus endlich vielen von ihnen ergeben. Ferner beweist Siegel in [1] und [2], daß sich unter der Voraussetzung von (4) und (5) sämtliche Ungleichungen (6) aus endlich vielen dieser Ungleichungen ergeben. Für beliebiges n bleibt jedoch die Frage offen, welche Ungleichungen wirklich notwendig sind. Dabei bedeutet die Notwendigkeit einer Ungleichung die Existenz eines Punktes aus $\mathfrak{H}$, in welchem diese eine Ungleichung nicht erfüllt ist, die übrigen Ungleichungen aber alle erfüllt sind. Für $n = 1$ kennt man diese notwendigen Ungleichungen; ist nämlich $z = x + iy$ die eine komplexe Veränderliche im Falle $n = 1$, so sind es die drei Ungleichungen

$$\frac{1}{2} \geqq x \geqq -\frac{1}{2}, \quad |z| \geqq 1.$$

Dieses Ergebnis wurde zuerst von R. Dedekind in [5] veröffentlicht. Es ist

das Ziel der vorliegenden Arbeit, die Frage nach den notwendigen Ungleichungen im Falle $n = 2$ zu beantworten. Setzt man in diesem Falle

$$Z = \begin{pmatrix} z_1 & z_3 \\ z_3 & z_2 \end{pmatrix} = \begin{pmatrix} x_1 & x_3 \\ x_3 & x_2 \end{pmatrix} + i \begin{pmatrix} y_1 & y_3 \\ y_3 & y_2 \end{pmatrix},$$

so liefert (4) die sechs Ungleichungen

(7) $$\frac{1}{2} \geqq x_1, x_2, x_3 \geqq -\frac{1}{2}.$$

Sämtliche Minkowskischen Reduktionsbedingungen sind bekanntlich, wie u. a. in [3] gezeigt wird, eine Folge der drei Ungleichungen

(8) $$y_2 \geqq y_1 \geqq 2 y_3 \geqq 0 .$$

Im ersten Paragraphen der vorliegenden Arbeit wird der folgende Satz bewiesen:

Satz 1: Unter der Voraussetzung von (7) *und* (8) *ergeben sich im Falle* $n = 2$ *alle Ungleichungen* (6) *bereits aus 19 dieser Ungleichungen, nämlich aus vier Ungleichungen, in denen Rang* $C = 1$ *ist:*

(9) $$|z_1| \geqq 1, \quad |z_2| \geqq 1, \quad |z_1 + z_2 - 2 z_3 + e| \geqq 1 \qquad (e = \pm 1),$$

und aus 15 Ungleichungen, in denen Rang $C = 2$ *ist:*

(10) $$\operatorname{abs}(Z + S) \geqq 1,$$

wobei für S *die 15 Matrizen*

$$\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} e & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & e \end{pmatrix}, \begin{pmatrix} e & 0 \\ 0 & e \end{pmatrix}, \begin{pmatrix} e & 0 \\ 0 & -e \end{pmatrix}, \begin{pmatrix} 0 & e \\ e & 0 \end{pmatrix}, \begin{pmatrix} e & e \\ e & 0 \end{pmatrix}, \begin{pmatrix} 0 & e \\ e & e \end{pmatrix}$$

einzusetzen sind.

Im zweiten Paragraphen wird die Notwendigkeit der 28 Ungleichungen (7), (8), (9), (10) für die Definition von $\mathfrak{F}$ bewiesen. Wir formulieren dieses Ergebnis als

Satz 2: Zu jeder der 28 Ungleichungen (7), (8), (9), (10) *gibt es einen Punkt von* $\mathfrak{H}$, *in welchem diese Ungleichung nicht erfüllt ist, die übrigen 27 Ungleichungen aber erfüllt sind. Geometrisch ausgedrückt: Von den fünfdimensionalen Flächen, die man erhält, wenn in den 28 Ungleichungen speziell das Gleichheitszeichen betrachtet wird, gehören fünfdimensionale Stücke zum Rande von* $\mathfrak{F}$.

§ 1. Überflüssige Ungleichungen

Mit C, D ist auch $-C, D$ teilerfremd und symmetrisch. Mit jedem Punkt $Z = X + iY$ von $\mathfrak{H}$ gehört auch der Punkt $-\overline{Z} = -X + iY$ zu $\mathfrak{H}$. Es gilt stets

(11) $$\operatorname{abs}(CZ + D) = \operatorname{abs}(-C(-\overline{Z}) + D).$$

Daraus folgt einmal, daß mit Z auch $-\overline{Z}$ zum Fundamentalbereich $\mathfrak{F}$ gehört, weil nämlich mit C, D auch $-C, D$ alle teilerfremden symmetrischen Matrizenpaare durchläuft und auf Grund von (11) die Ungleichungen $\operatorname{abs}(CZ + D) \geqq 1$ und $\operatorname{abs}(-C(-\overline{Z}) + D) \geqq 1$ gleichbedeutend sind, und weil die Ungleichungen (4) und (5) beim Übergang von Z zu $-\overline{Z}$ erhalten bleiben. Es ergibt sich ferner

Lemma 1: Die Ungleichungen $\mathrm{abs}(CZ + D) \geqq 1$ *und* $\mathrm{abs}(-CZ + D) \geqq 1$ *sind entweder beide notwendig oder beide überflüssig für die Definition von* $\mathfrak{F}$.

Die Ungleichung $\mathrm{abs}(CZ + D) \geqq 1$ ist nämlich genau dann notwendig, wenn es eine Stelle Z_0 gibt, für die $\mathrm{abs}(CZ_0 + D) < 1$ ist, sonst aber alle Ungleichungen (4), (5), (6) erfüllt sind. Für die Stelle $-\overline{Z}_0$ gilt dann $\mathrm{abs}(-C(-\overline{Z}_0) + D) < 1$, während alle anderen Ungleichungen für $-\overline{Z}_0$ erfüllt sind.

Die letzten Betrachtungen galten für beliebiges n. Jetzt sei $n = 2$. Unter den Ungleichungen (6) kommen dann insbesondere die folgenden beiden vor:

$$|z_1| \geqq 1, \quad |z_2| \geqq 1. \tag{12}$$

Man erhält sie für die beiden teilerfremden symmetrischen Paare

$$C, D = \begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}, \begin{pmatrix}0 & 0\\ 0 & 1\end{pmatrix}; \quad C, D = \begin{pmatrix}0 & 0\\ 0 & 1\end{pmatrix}, \begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}.$$

Durch die Ungleichungen (7), (8), (12) wird ein Bereich $\mathfrak{B}$ definiert, der $\mathfrak{F}$ enthält, und in welchem insbesondere

$$y_1^2 \geqq \frac{3}{4}, \quad y_2^2 \geqq \frac{3}{4}, \quad \det Y \geqq \frac{3}{4}\, y_1^2 \geqq \frac{9}{16} \tag{13}$$

gilt. Dieser Bereich $\mathfrak{B}$ wird im folgenden wiederholt benutzt werden.

Es sei C, D ein teilerfremdes symmetrisches Paar mit *Rang* $C = 1$. In diesem Falle gibt es zwei unimodulare Matrizen U und V, so daß

$$UCV = \begin{pmatrix}c & 0\\ 0 & 0\end{pmatrix} \qquad (c > 0)$$

wird. Setzt man

$$UDV'^{-1} = \begin{pmatrix}d & d_1\\ d_2 & d_3\end{pmatrix},$$

so folgt aus $CD' = DC'$ die Bedingung $d_2 = 0$. Aus der Teilerfremdheit von C, D ergibt sich die Teilerfremdheit von c, d und $d_3 = \pm 1$. Ersetzt man noch U durch

$$\begin{pmatrix}1 & d_1\\ 0 & d_3\end{pmatrix} U,$$

so wird

$$UCV = \begin{pmatrix}c & 0\\ 0 & 0\end{pmatrix}, \quad UDV'^{-1} = \begin{pmatrix}d & 0\\ 0 & 1\end{pmatrix}. \tag{14}$$

Hieraus folgt, wenn die erste Zeile von V^{-1} mit (pq) bezeichnet wird:

$$\mathrm{abs}(CZ + D) = |c(z_1 p^2 + z_2 q^2 + 2 z_3 pq) + d|. \tag{15}$$

Dabei sind also p und q teilerfremd, und wir können offenbar noch $p > 0$ oder aber $p = 0, q = 1$ voraussetzen. Sind umgekehrt die vier Zahlen c, d, p, q mit

$$(c, d) = 1; \quad (p, q) = 1; \quad c > 0; \quad p > 0 \text{ oder } p = 0, q = 1 \tag{16}$$

vorgegeben, so ergänze man die Zeile (pq) zu einer unimodularen Matrix V^{-1}. Das Matrizenpaar

$$C = \begin{pmatrix}c & 0\\ 0 & 0\end{pmatrix} V^{-1}, \quad D = \begin{pmatrix}d & 0\\ 0 & 1\end{pmatrix} V' \tag{17}$$

ist dann teilerfremd und symmetrisch, und es gilt *Rang* $C = 1$.

Wir zeigen noch, daß die Zuordnung zwischen den Klassen assoziierter Paare C, D mit *Rang* $C = 1$ und den vier ganzen Zahlen c, d, p, q mit den Eigenschaften (16) umkehrbar eindeutig ist. Ist neben V^{-1} auch V_1^{-1} eine unimodulare Matrix mit (pq) als erster Zeile, und wird analog zu (17) noch

$$C_1 = \begin{pmatrix} c & 0 \\ 0 & 0 \end{pmatrix} V_1^{-1}, \quad D_1 = \begin{pmatrix} d & 0 \\ 0 & 1 \end{pmatrix} V_1' \tag{18}$$

gesetzt, so folgen aus (17) und (18) die Gleichungen

$$C_1 = U_1 C, \quad D_1 = U_1 D,$$

wo U_1 die unimodulare Matrix

$$U_1 = \begin{pmatrix} d & 0 \\ 0 & 1 \end{pmatrix} V_1' \, V'^{-1} \begin{pmatrix} d^{-1} & 0 \\ 0 & 1 \end{pmatrix}$$

ist. Also sind die Paare C_1, D_1 und C, D assoziiert. Durch Vorgabe von c, d, p, q ist somit die Klasse der zu C, D assoziierten Paare eindeutig bestimmt. Es sei umgekehrt C, D vorgegeben, und es gelte neben (14) auch entsprechend

$$U_0 C V_0 = \begin{pmatrix} c_0 & 0 \\ 0 & 0 \end{pmatrix}, \quad U_0 D V_0'^{-1} = \begin{pmatrix} d_0 & 0 \\ 0 & 1 \end{pmatrix},$$

wo die erste Zeile von V_0^{-1} mit $(p_0 q_0)$ bezeichnet werde und für die vier ganzen Zahlen c_0, d_0, p_0, q_0 entsprechende Voraussetzungen gelten wie für c, d, p, q. Aus

$$U_0 U^{-1} \begin{pmatrix} c & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} c_0 & 0 \\ 0 & 0 \end{pmatrix} V_0^{-1} V$$

folgt dann wegen $c > 0$ und $c_0 > 0$ die Gleichung $c = c_0$. Außerdem ergibt sich

$$V_0^{-1} V = \begin{pmatrix} \pm 1 & 0 \\ * & \pm 1 \end{pmatrix}.$$

Das bedeutet wegen der Voraussetzungen über p, q und p_0, q_0 insbesondere, daß die ersten Zeilen von V_0^{-1} und V^{-1} übereinstimmen. Beachten wir dann noch

$$U U_0^{-1} \begin{pmatrix} d_0 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} d & 0 \\ 0 & 1 \end{pmatrix} V' V_0'^{-1},$$

so folgt $d_0 = d$.

Nach (15) ist also jeder Ungleichung

$$\operatorname{abs}(CZ + D) \geqq 1 \qquad (\textit{Rang}\ C = 1)$$

umkehrbar eindeutig eine Ungleichung der Form

$$|c(z_1 p^2 + z_2 q^2 + 2 z_3 pq) + d| \geqq 1$$

zugeordnet, wo die ganzen Zahlen c, d, p, q den Bedingungen (16) genügen.

Die rechte Seite von (15) schätzen wir durch den Absolutbetrag des Imaginärteils ab:

$$\operatorname{abs}(CZ + D) \geqq c(y_1 p^2 + y_2 q^2 + 2 y_3 pq). \tag{19}$$

Aus $|pq| \leqq \frac{1}{2}(p^2 + q^2)$ ergibt sich dann nach (8) und (13) für den ganzen

Bereich $\mathfrak{B}$ die Abschätzung

$$\text{abs}(CZ + D) \geqq \frac{1}{4}\sqrt{3}\,c(p^2+q^2) \geqq \frac{1}{4}\sqrt{3}\,(p^2+q^2). \tag{20}$$

Wenn $pq \neq 0$ ist, folgert man hieraus

$$\text{abs}(CZ+D) \geqq \frac{1}{2}\sqrt{3c}\,, \tag{21}$$

und dies gilt, wie man direkt aus (19) sieht, auch noch, falls $p = 0$ oder $q = 0$ ist. Wegen $(p, q) = 1$ können p und q nicht gleichzeitig verschwinden. Aufgrund der beiden letzten Ungleichungen (20) und (21) gilt im ganzen Bereich $\mathfrak{B}$ stets $\text{abs}(CZ + D) > 1$, falls nur $c \geqq 2$ oder $p^2 + q^2 \geqq 4$ ist. Daher sind von den Ungleichungen (6) mit ausgeartetem C höchstens diejenigen mit $c = 1$ und $p^2 + q^2 < 4$ notwendig (die beiden Ungleichungen (12) kommen unter diesen vor). Wegen (16) brauchen wir nur die Fälle

$$c = 1\,, \quad p, q = 1, 0;\ 0, 1;\ 1, 1;\ 1, -1$$

zu untersuchen. Der Fall $p, q = 1, 0$ liefert nach (15) die Beziehung

$$\text{abs}(CZ + D)^2 = |z_1 + d|^2 = (x_1 + d)^2 + y_1^2\,.$$

Wenn $d \neq 0$ ist, folgt im Bereich $\mathfrak{B}$ die Ungleichung $|x_1 + d| \geqq \frac{1}{2}$ und daher wegen $y_1^2 \geqq \frac{3}{4}$ auch $\text{abs}(CZ + D) \geqq 1$. Somit ist also höchstens die Ungleichung mit $d = 0$ notwendig, das ist die erste der Ungleichungen (12). Entsprechend zeigt man, daß sich im Falle $p, q = 0, 1$ höchstens die zweite der Ungleichungen (12) als notwendig erweist. Für $p, q = 1, 1$ gilt nach (19), (8), (13) im Bereich $\mathfrak{B}$ die Abschätzung

$$\text{abs}(CZ + D) \geqq y_1 + y_2 + 2y_3 \geqq \sqrt{3} > 1\,.$$

Die Ungleichungen mit $p, q = 1, 1$ sind daher sämtlich überflüssig. Ist schließlich $p, q = 1, -1$, so gilt nach (15) die Gleichung

$$\text{abs}(CZ + D)^2 = (x_1 + x_2 - 2x_3 + d)^2 + (y_1 + y_2 - 2y_3)^2\,.$$

Für $|d| \geqq 3$ ist im Bereich $\mathfrak{B}$ stets $|x_1 + x_2 - 2x_3 + d| \geqq 1$, also auch $\text{abs}(CZ + D) \geqq 1$. Daher sind in diesem Falle höchstens die Ungleichungen

$$|z_1 + z_2 - 2z_3 + d| \geqq 1 \qquad (-2 \leqq d \leqq 2) \tag{22}$$

notwendig. Von diesen Ungleichungen werden sich im nächsten Lemma noch diejenigen mit $d = 0$ und $d = \pm 2$ als überflüssig erweisen, so daß tatsächlich von allen Ungleichungen (6) mit *Rang* $C = 1$ höchstens die vier Ungleichungen (9) notwendig sind.

Wegen $y_1 \geqq 2y_3$ ist $y_1 + y_2 - 2y_3 \geqq y_2$. Daher sind die Ungleichungen (22) sicherlich für $y_2 \geqq 1$ erfüllt. Es sei $\mathfrak{B}_0$ derjenige Teil von $\mathfrak{B}$, in welchem $y_2 \leqq 1$ ist. Von den Ungleichungen (22) mit $d = \pm 2$ brauchte nun aufgrund von Lemma 1 eigentlich nur etwa diejenige mit $d = 2$ untersucht zu werden. Es ist jedoch nicht mühsamer, beide Fälle zugleich zu behandeln. Die Ungleichungen (22) mit $d = -2e$ $(e = \pm 1)$ sind nämlich sicher dann erfüllt, wenn

$|x_1 + x_2 - 2x_3 - 2e| \geqq \frac{1}{2}$ ist. Ist dagegen $|x_1 + x_2 - 2x_3 - 2e| \leqq \frac{1}{2}$, so folgt nach (7) in $\mathfrak{B}$ für $e = 1$ die Bedingung

$$-\frac{1}{4} \geqq x_3 \geqq -\frac{1}{2}, \tag{23}$$

und für $e = -1$ die Bedingung

$$\frac{1}{2} \geqq x_3 \geqq \frac{1}{4}. \tag{24}$$

Es sei $\mathfrak{B}_1$ derjenige Teil von $\mathfrak{B}_0$, in welchem (23) gilt, und $\mathfrak{B}_{-1}$ derjenige Teil von $\mathfrak{B}_0$, in welchem (24) gilt.

Lemma 2: Im Bereich $\mathfrak{B}_e$ $(e = -1, 0, 1)$ *ist die Ungleichung*

$$|z_1 + z_2 - 2z_3 - 2e| \geqq 1$$

eine Folge von $\operatorname{abs}(Z + eS) \geqq 1$ *mit* $S = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$.

Beweis: Die Differenz

$$F_e = |z_1 + z_2 - 2z_3 - 2e|^2 - \operatorname{abs}(Z + eS)^2$$

geht aus der Differenz

$$F_0 = F = |z_1 + z_2 - 2z_3|^2 - \operatorname{abs} Z^2$$

dadurch hervor, daß x_3 durch $e + x_3$ ersetzt wird. Im Falle $e = 1$ genügt x_3 der Bedingung (23). In diesem Falle gilt für $e + x_3$ die Ungleichung $\frac{3}{4} \geqq e + x_3 \geqq \frac{1}{2}$. Im Falle $e = -1$ genügt x_3 der Bedingung (24). Daher gilt in diesem Falle für $e + x_3$ die Ungleichung $-\frac{1}{2} \geqq e + x_3 \geqq -\frac{3}{4}$. Für den Beweis von Lemma 2 genügt es deswegen zu zeigen, daß $F \geqq 0$ ist in demjenigen Bereich $\mathfrak{B}^*$, der aus $\mathfrak{B}_0$ hervorgeht, indem die Bedingung $|x_3| \leqq \frac{1}{2}$ durch $|x_3| \leqq \frac{3}{4}$ ersetzt wird.

Weil wegen (12) für die zweite Ableitung

$$F_{x_1 x_1} = 2 - 2(x_2^2 + y_2^2) \leqq 0$$

gilt, ist F als Funktion von x_1 nach unten konvex. Wenn daher $F \geqq 0$ für $x_1 = \pm\frac{1}{2}$ erfüllt ist, dann auch für $|x_1| \leqq \frac{1}{2}$. Da außerdem auch

$$F_{x_2 x_2} = 2 - 2(x_1^2 + y_1^2) \leqq 0$$

ist und F ungeändert bleibt, wenn Z durch $-\overline{Z}$, also X durch $-X$ ersetzt wird, genügt es, die beiden Fälle

$$x_1 = \frac{1}{2}, \quad x_2 = \frac{1}{2} \quad \text{und} \quad x_1 = \frac{1}{2}, \quad x_2 = -\frac{1}{2}$$

zu betrachten. Hierbei wird auch noch benutzt, daß $\mathfrak{B}^*$ bei der Substitution $Z \to -\overline{Z}$ in sich übergeht. Setzen wir im ersten Falle $x = (1 - 2x_3)^2$, $e = -1$

und im zweiten Falle $x = 4x_3^2$, $e = 1$, so wird in beiden Fällen

$$F = x + (y_1 + y_2 - 2y_3)^2 - \left(\det Y + x_3^2 + \frac{e}{4}\right)^2 - \frac{1}{4}(y_2 - e y_1 - 4x_3 y_3)^2 .$$

Wir zeigen zuerst, daß $F_{y_3} \leqq 0$ ist. Setzen wir

$$H = 4y_3\left(\det Y + \frac{e}{4}\right), \qquad K = 2x_3(y_2 - e y_1 - 2x_3 y_3),$$

so folgt in ganz $\mathfrak{B}^*$ für die betrachtete Ableitung:

$$F_{y_3} = -4(y_1 + y_2 - 2y_3) + H + K \leqq -4y_2 + H + K .$$

Für $|x_3| \leqq \frac{1}{2}$ gilt $K \leqq y_2 - e y_1$, außerdem $H \leqq 2y_1\left(1 + \frac{e}{4}\right)$. Daher

$$F_{y_3} \leqq -4y_2 + 2y_1\left(1 + \frac{e}{4}\right) + y_2 - e y_1 < 0 .$$

Für $\frac{1}{2} \leqq |x_3| \leqq \frac{3}{4}$ schätzen wir etwas anders ab:

$$H \leqq 4y_3\left(1 + \frac{e}{4}\right), \quad K \leqq \frac{3}{2}(y_2 - e y_1) - y_3 .$$

Damit ergibt sich

$$F_{y_3} \leqq -4y_2 + y_3(3 + e) + \frac{3}{2}(y_2 - e y_1) \leqq 0 .$$

Also ist F als Funktion von y_3 monoton fallend und nimmt daher den minimalen Wert für den größtmöglichen Wert von y_3, d. h. für $y_3 = \frac{1}{2} y_1$ an. Dafür wird

$$F = x + y_2^2 - \left(y_1 y_2 - \frac{1}{4} y_1^2 + x_3^2 + \frac{e}{4}\right)^2 - \frac{1}{4}(y_2 - e y_1 - 2x_3 y_1)^2 .$$

Wir zeigen jetzt, daß auch $F_{y_1} \leqq 0$ ist. Setzen wir

$$L = \frac{1}{2}(e + 2x_3)(y_2 - e y_1 - 2x_3 y_1),$$

so wird

$$F_{y_1} = -(2y_2 - y_1)\left(y_1 y_2 - \frac{1}{4} y_1^2 + x_3^2 + \frac{e}{4}\right) + L \leqq -y_2\left(\frac{1}{2} + \frac{e}{4}\right) + L .$$

Für $e = -1$ und $\frac{1}{2} \geqq x_3 \geqq -\frac{3}{4}$ gilt $L \leqq 0$, für $e = -1$ und $\frac{3}{4} \geqq x_3 \geqq \frac{1}{2}$ jedenfalls $L \leqq \frac{1}{4} y_2$, also für $e = -1$ stets $F_{y_1} \leqq 0$. Für $e = 1$ und $\frac{3}{4} \geqq |x_3| \geqq \frac{1}{2}$ wird $L \leqq 0$, für $e = 1$ und $\frac{1}{2} \geqq x_3 \geqq 0$ jedenfalls $L \leqq y_2 - y_1$, für $e = 1$ und $0 \geqq x_3 \geqq -\frac{1}{2}$ schließlich $L \leqq \frac{1}{2} y_2$. Beachten wir $1 \geqq y_2 \geqq y_1 \geqq \frac{1}{2}\sqrt{3}$, so folgt auch für $e = 1$ stets $F_{y_1} < 0$. Somit nimmt F den minimalen Wert für den größtmöglichen Wert von y_1, d. h. für $y_1 = y_2$ an. Dafür wird

$$F = x + y_2^2 - \left(\frac{3}{4} y_2^2 + x_3^2 + \frac{e}{4}\right)^2 - \frac{1}{4} y_2^2(1 - e - 2x_3)^2 .$$

Wir zeigen endlich auch noch, daß $F_{y_2} \leqq 0$ ist:

$$F_{y_2} = 2y_2 - 3y_2\left(\frac{3}{4} y_2^2 + x_3^2 + \frac{e}{4}\right) - \frac{1}{2} y_2(1 - e - 2x_3)^2 .$$

Für $e = 1$ folgt sofort $F_{y_2} < 0$. Für $e = -1$ wird

$$y_2^{-1} F_{y_2} = -3\left(\frac{3}{4} y_2^2 - \frac{1}{4}\right) - 5x_3^2 + 4x_3.$$

Weil aber für alle x_3 die Ungleichung $4x_3 - 5x_3^2 \leq \frac{4}{5}$ gilt, folgt auch in diesem Falle $F_{y_2} < 0$. Also nimmt F den minimalen Wert für $y_2 = 1$ an. Im Falle $e = 1$ erhalten wir damit

$$F = 4x_3^2 + 1 - (1 + x_3^2)^2 - x_3^2 = x_3^2 - x_3^4 \geq 0,$$

weil ja jedenfalls $|x_3| \leq 1$ ist. Für $e = -1$ erhalten wir

$$F = (1 - 2x_3)^2 + 1 - \left(\frac{1}{2} + x_3^2\right)^2 - (1 - x_3)^2 = \frac{3}{4} - 2x_3 + 2x_3^2 - x_3^4.$$

Wir bilden die Ableitung

$$F_{x_3} = -2 + 4x_3(1 - x_3^2)$$

und wollen $F_{x_3} \leq 0$ beweisen. Dies ist sicher für $0 \geq x_3 \geq -\frac{3}{4}$ richtig. Für jedes Intervall $0 \leq a \leq x_3 \leq b \leq 1$ gilt ferner die Abschätzung

$$4x_3(1 - x_3^2) \leq 4b(1 - a^2).$$

Wenden wir das in den drei Fällen $a, b = 0, \frac{1}{2}$; $\frac{1}{2}, \frac{2}{3}$; $\frac{2}{3}, \frac{3}{4}$ an, so finden wir stets $F_{x_3} \leq 0$. Also nimmt F den minimalen Wert für $x_3 = \frac{3}{4}$ an. Dafür wird

$$F = \frac{3}{4} - \frac{3}{2} + \frac{9}{8} - \frac{81}{256} > 0.$$

Damit ist Lemma 2 bewiesen.

Wir wenden uns nun dem Fall *Rang* $C = 2$ zu. In diesem Falle können wir

$$\operatorname{abs}(CZ + D) = \operatorname{abs} C \cdot \operatorname{abs}(Z + C^{-1}D)$$

schreiben. Setzen wir $R = X + C^{-1}D$, so ist R symmetrisch. Weil Y positiv definit ist, gibt es daher eine reelle Matrix G mit

$$Y = GG', \quad R = GHG', \quad H = \begin{pmatrix} h_1 & 0 \\ 0 & h_2 \end{pmatrix}. \tag{25}$$

Es wird

$$R + iY = G(H + iE)G'$$

und daher

$$\operatorname{abs}(CZ + D)^2 = \det C^2 \cdot \det Y^2 \cdot (h_1^2 + 1)(h_2^2 + 1). \tag{26}$$

Für $\operatorname{abs} C \geq 2$ folgt im ganzen Bereich $\mathfrak{B}$ wegen (13) die Abschätzung

$$\operatorname{abs}(CZ + D) \geq 2 \cdot \frac{9}{16} > 1.$$

Es sind daher höchstens die Ungleichungen mit $\det C = \pm 1$ notwendig. Für diese ist also C unimodular; indem eventuell noch C, D durch ein assoziiertes

Paar ersetzt wird, kann $C, D = E, S$ angenommen werden, wo S eine symmetrische Matrix mit ganzzahligen Elementen ist. Es handelt sich dann darum, noch die Ungleichungen

$$\text{abs}(Z + S) \geqq 1 \tag{27}$$

zu untersuchen. Setzen wir

$$G = \begin{pmatrix} g_1 & g_3 \\ g_4 & g_2 \end{pmatrix}, \quad S = \begin{pmatrix} s_1 & s_3 \\ s_3 & s_2 \end{pmatrix}$$

und beachten $R = X + S$, so schreibt sich (25) folgendermaßen in Komponentenform:

$$y_1 = g_1^2 + g_3^2, \quad y_2 = g_4^2 + g_2^2, \quad y_3 = g_1 g_4 + g_2 g_3$$

und

$$x_1 + s_1 = h_1 g_1^2 + h_2 g_3^2, \quad x_2 + s_2 = h_1 g_4^2 + h_2 g_2^2, \quad x_3 + s_3 = h_1 g_1 g_4 + h_2 g_2 g_3.$$

Setzen wir ferner $Max\,(|h_1|, |h_2|) = h$ und beachten

$$|g_1 g_4| + |g_2 g_3| \leqq \frac{1}{2}(g_1^2 + g_4^2 + g_2^2 + g_3^2) = \frac{1}{2}(y_1 + y_2) \leqq y_2,$$

so folgt

$$|x_k + s_k| \leqq y_2 h \qquad (k = 1, 2, 3).$$

Wäre nun für einen der Werte $k = 1, 2, 3$ die Ungleichung $|s_k| \geqq 2$ erfüllt, so folgte wegen $|x_k| \leqq \frac{1}{2}$ für h die Abschätzung $y_2 h \geqq \frac{3}{2}$, also nach (26) die Abschätzung

$$\text{abs}(Z + S)^2 \geqq \det Y^2 \left(1 + \frac{9}{4} y_2^{-2}\right) = (y_1 - y_3^2 y_2^{-1})^2 \left(y_2^2 + \frac{9}{4}\right) > 1.$$

Daher sind von den Ungleichungen (27) höchstens diejenigen notwendig, in denen $|s_1|, |s_2|, |s_3| \leqq 1$ ist. Das sind außer den Ungleichungen (10) noch genau diejenigen mit

$$S = \begin{pmatrix} e & e \\ e & -e \end{pmatrix}, \begin{pmatrix} e & -e \\ -e & -e \end{pmatrix}, \tag{28}$$

$$S = \begin{pmatrix} e & -e \\ -e & e \end{pmatrix}, \tag{29}$$

$$S = \begin{pmatrix} e & -e \\ -e & 0 \end{pmatrix}, \begin{pmatrix} 0 & -e \\ -e & e \end{pmatrix}, \tag{30}$$

$$S = \begin{pmatrix} e & e \\ e & e \end{pmatrix}, \tag{31}$$

wo für e überall die beiden Möglichkeiten $e = \pm 1$ einzutragen sind. Aufgrund des folgenden Lemmas erweisen sich die Ungleichungen (27) in den Fällen (28), (29), (30) als überflüssig.

Lemma 3: In den Fällen (28), (29), (30) *gilt im Bereich* $\mathfrak{B}$ *überall* $\text{abs}(Z + S) \geqq 1$.

Beweis: Da mit $Z = X + iY$ auch $-\bar{Z} = -X + iY$ zu $\mathfrak{B}$ gehört, genügt es, in jedem Falle nur $e = 1$ zu betrachten. In dem Falle (28) gilt dann wegen $|x_k| \leqq \frac{1}{2}$ $(k = 1, 2, 3)$ die Ungleichung

$$\det(X + S) = (x_1 + 1)(x_2 - 1) - (x_3 \pm 1)^2 \leqq -\frac{1}{2}.$$

Wenn wir daher den Absolutbetrag $\operatorname{abs}(Z + S)$ durch den Absolutbetrag des Realteils von $\det(Z + S)$ abschätzen und noch (13) benutzen, folgt

$$\operatorname{abs}(Z + S) \geqq |\det(X + S) - \det Y| \geqq \frac{1}{2} + \frac{9}{16} > 1 .$$

Im Falle (29) untersuchen wir den Ausdruck

$$F = \operatorname{abs} Z^2 = (y_1 y_2 - y_3^2 - x_1 x_2 + x_3^2)^2 + (x_1 y_2 + x_2 y_1 - 2 x_3 y_3)^2 \tag{32}$$

für $x_1, x_2 \geqq \frac{1}{2}$ und $-\frac{1}{2} \geqq x_3$. Zunächst gilt wegen $y_3 \geqq 0$ und $y_2 \geqq y_1 \geqq \frac{1}{2}\sqrt{3}$ die Abschätzung

$$F \geqq (x_1 y_2 + x_2 y_1)^2 \geqq \frac{3}{4}(x_1 + x_2)^2 .$$

Für $x_1 + x_2 \geqq \frac{2}{3}\sqrt{3}$ folgt daher $F \geqq 1$. Ist dagegen $x_1 + x_2 \leqq \frac{2}{3}\sqrt{3}$, so folgt $x_1 x_2 \leqq \frac{1}{3}\sqrt{3}$, weil das geometrische Mittel zweier positiver Zahlen höchstens gleich dem arithmetischen ist. Deswegen gilt für den Realteil von $-\det Z$ die Abschätzung

$$y_1 y_2 - y_3^2 - x_1 x_2 + x_3^2 > \frac{3}{4} - x_1 x_2 > 0 ,$$

und daher

$$F \geqq \left(\frac{3}{4} - x_1 x_2\right)^2 + \frac{3}{4}(x_1 + x_2)^2 = \frac{9}{16} + x_1^2 x_2^2 + \frac{3}{4}(x_1^2 + x_2^2) \geqq 1 .$$

Für die Behandlung des Falles (30) ersetzen wir $\mathfrak{B}$ durch den Bereich $\mathfrak{B}'$, welcher durch die Ungleichungen $|x_k| \leqq \frac{1}{2}$ $(k = 1, 2, 3)$ und

$$y_3 \geqq 0 , \quad y_1, y_2 \geqq 2 y_3 , \quad y_1, y_2 \geqq \frac{1}{2}\sqrt{3} \tag{33}$$

definiert ist. Der Bereich $\mathfrak{B}'$ enthält $\mathfrak{B}$ und hat ebenfalls die Eigenschaft, mit $Z = X + iY$ auch den Punkt $-\overline{Z} = -X + iY$ zu enthalten. Ferner geht aber $\mathfrak{B}'$ auch in sich über, wenn die beiden Variablen z_1, z_2 miteinander vertauscht werden. Wenn wir daher die Ungleichung $\operatorname{abs}(Z + S) \geqq 1$ in $\mathfrak{B}'$ für den Fall $S = \begin{pmatrix} 1 & -1 \\ -1 & 0 \end{pmatrix}$ bewiesen haben, dann gilt sie in $\mathfrak{B}'$ auch für den Fall $S = \begin{pmatrix} 0 & -1 \\ -1 & 1 \end{pmatrix}$. Also genügt es zu zeigen, daß der Ausdruck (32) unter der Voraussetzung von (33) und

$$\frac{3}{2} \geqq x_1 \geqq \frac{1}{2} , \quad \frac{1}{2} \geqq x_2 \geqq -\frac{1}{2} , \quad -\frac{1}{2} \geqq x_3 \geqq -\frac{3}{2} \tag{34}$$

stets der Ungleichung $F \geqq 1$ genügt. Zunächst ergeben sich aus (33) und (34) für die Ableitungen von F nach y_1 und y_2 die Abschätzungen

$$F_{y_1} = 2 y_2 (y_1 y_2 - y_3^2 + x_3^2) + 2 y_1 x_2^2 - 4 x_2 x_3 y_3 > 0 ,$$
$$F_{y_2} = 2 y_1 (y_1 y_2 - y_3^2 + x_3^2) + 2 y_2 x_1^2 - 4 x_1 x_3 y_3 > 0 .$$

Der Ausdruck F nimmt also den minimalen Wert für die kleinstmöglichen

Werte von y_1 und y_2 an. Diese sind nach (33) entweder

(35) $$y_1 = y_2 = 2y_3 \quad \textit{falls} \quad y_3 \geqq \frac{1}{4}\sqrt{3}$$

oder

(36) $$y_1 = y_2 = \frac{1}{2}\sqrt{3} \quad \textit{falls} \quad y_3 \leqq \frac{1}{4}\sqrt{3}\,.$$

Wird F als Funktion von x_3 betrachtet, so nimmt F in den beiden Fällen (35) und (36), wie man leicht aus (32) sieht, den minimalen Wert für $x_3 = -\frac{1}{2}$ an. Im Falle (35) ergibt sich dafür

(37) $$F = \left(\frac{1}{4} + 3y_3^2 - x_1x_2\right)^2 + y_3^2(2x_1 + 2x_2 + 1)^2\,.$$

Im Falle (36) erhalten wir

(38) $$F = (1 - y_3^2 - x_1x_2)^2 + \frac{3}{4}\left(x_1 + x_2 + \frac{2}{3}\sqrt{3}\,y_3\right)^2.$$

Wir behandeln zuerst (37). In diesem Falle ist F eine monoton wachsende Funktion von y_3 und nimmt daher den minimalen Wert für $y_3 = \frac{1}{4}\sqrt{3}$ an; dafür ist

$$F \geqq \left(\frac{3}{4} - x_1x_2\right)^2 + \frac{3}{4}\left(x_1 + x_2 + \frac{1}{2}\right)^2 \geqq \frac{3}{4}(1 + x_1^2 + x_2^2 + x_1 + x_2).$$

Beachten wir $x_1 \geqq \frac{1}{2}$, und daß für alle reellen x_2 die Ungleichung $x_2^2 + x_2 \geqq -\frac{1}{4}$ gilt, so folgt $F > 1$. Für die Behandlung der Gleichung (38) wird F nach x_1 differenziert:

$$F_{x_1} = y_3(\sqrt{3} + 2x_2y_3) + 2x_1x_2^2 + \frac{1}{2}(3x_1 - x_2) > 0\,.$$

Also nimmt F den minimalen Wert für $x_1 = \frac{1}{2}$ an. Beachten wir noch $y_3^2 \leqq \frac{1}{3}\sqrt{3}\,y_3$, so folgt

$$F \geqq \left(1 - \frac{1}{2}x_2 - y_3^2\right)^2 + \frac{3}{4}\left(\frac{1}{2} + x_2 + 2y_3^2\right)^2.$$

Setzen wir jetzt $\frac{1}{2}x_2 + y_3^2 = x$, so wird

$$F \geqq (1 - x)^2 + \frac{3}{4}\left(\frac{1}{2} + 2x\right)^2 = 1 + \frac{3}{16} + 4x^2 - \frac{1}{2}x > 1\,.$$

Damit ist Lemma 3 bewiesen.

Um nun noch den Fall (31) zu erledigen, beachten wir, daß die Ungleichungen (27) jedenfalls für $y_1 \geqq \frac{2}{3}\sqrt{3}$ erfüllt sind: Nach (26) gilt nämlich

$$\operatorname{abs}(Z + S) \geqq \det Y \geqq \frac{3}{4}y_1^2\,.$$

Es sei $\mathfrak{B}_2$ derjenige Teil von $\mathfrak{B}$, in welchem $y_1 \leqq \frac{2}{3}\sqrt{3}$ ist.

Lemma 4: Im Bereich $\mathfrak{B}_2$ *ist die Ungleichung* $\operatorname{abs}(Z+S) \geqq 1$ *mit* $S = \begin{pmatrix} e & e \\ e & e \end{pmatrix}$ $(e = \pm 1)$ *eine Folge der Ungleichung* $\operatorname{abs} Z \geqq 1$.

Beweis: Es genügt wieder, nur den Fall $e = 1$ zu betrachten. Wir bilden die Differenz

$$F = \operatorname{abs}(Z+S)^2 - \operatorname{abs} Z^2 =$$
$$= |z_1 + z_2 - 2z_3|^2 - 2(x_1 + x_2 - 2x_3)(\det Y + x_3^2 - x_1 x_2) +$$
$$+ 2(y_1 + y_2 - 2y_3)(x_1 y_2 + x_2 y_1 - 2x_3 y_3)$$

und beweisen, daß in $\mathfrak{B}_2$ überall $F \geqq 0$ ist. Es wird

$$F_{y_2} = (2 + 4x_3)(y_1 - y_3) + (2 + 4x_1)(y_2 - y_3) \geqq 0\,.$$

Also nimmt F den minimalen Wert für den kleinstmöglichen Wert von y_2, d. h. für $y_2 = y_1$ an. Dafür ist

$$F = (x_1 + x_2 - 2x_3)^2 + 4(y_1 - y_3)^2 - 2(x_1 + x_2 - 2x_3)(y_1^2 - y_3^2 + x_3^2 - x_1 x_2) +$$
$$+ 4(y_1 - y_3)(y_1(x_1 + x_2) - 2x_3 y_3)\,.$$

Für die Ableitung nach y_3 ergibt sich

$$F_{y_3} = (y_1 - y_3)(-8 - 4(x_1 + x_2) - 8x_3) \leqq 0\,.$$

Also nimmt F den minimalen Wert für $y_3 = \frac{1}{2} y_1$ an. Dafür ist

(39) $$F = (x_1 + x_2 - 2x_3)^2 + y_1^2(1 + 2x_3) +$$
$$+ 2(x_1 + x_2 - 2x_3)\left(\frac{1}{4} y_1^2 - x_3^2 + x_1 x_2\right).$$

Wir beachten nun, daß im ganzen Bereich $\mathfrak{B}_2$ die Abschätzung

(40) $$\left|\frac{1}{4} y_1^2 - x_3^2 + x_1 x_2\right| \leqq \left|\frac{1}{4} y_1^2 - x_3^2\right| + \frac{1}{4} \leqq \frac{1}{4}(1 + y_1^2)$$

gilt und fragen danach, für welche positiven a die Ungleichung

(41) $$\frac{1}{4}(1 + y_1^2) \leqq a\, y_1$$

im ganzen Intervall $\frac{1}{2}\sqrt{3} \leqq y_1 \leqq \frac{2}{3}\sqrt{3}$ erfüllt ist. Dies ist genau für $a \geqq \frac{7}{24}\sqrt{3}$ der Fall. Für diese a können wir F nach (40) und (41) folgendermaßen abschätzen:

$$F \geqq (|x_1 + x_2 - 2x_3| - a\, y_1)^2 + y_1^2(1 + 2x_3 - a^2)\,.$$

Es folgt $F \geqq 0$, falls $1 - a^2 + 2x_3 \geqq 0$ ist. Wir wählen für a den kleinstmöglichen Wert und setzen für x_3 im folgenden die Ungleichung $1 - a^2 + 2x_3 \leqq 0$ oder

(42) $$-\frac{1}{2} + \frac{1}{2}a^2 \geqq x_3 \geqq -\frac{1}{2} \qquad \left(a = \frac{7}{24}\sqrt{3}\right)$$

voraus. Wenn nun

(43) $$x_1 + x_2 - 2x_3 \leqq 0$$

sein sollte, bilden wir in (39) die Ableitung nach x_3 und erhalten

$$F_{x_3} = -4(1 + x_3)(x_1 + x_2 - 2x_3) + y_1^2 + 4x_3^2 - 4x_1x_2\,.$$

Wegen (42) ist diese Ableitung positiv. Also nimmt F den minimalen Wert für den kleinstmöglichen Wert von x_3, das bedeutet nach (43) für $x_3 = \frac{1}{2}(x_1 + x_2) \geqq -\frac{1}{2}$ an. Dafür ist aber nach (39) sicher $F \geqq 0$. Ist dagegen

$$x_1 + x_2 - 2x_3 \geqq 0\,,$$

so schreiben wir

$$F = y_1^2(1 + 2x_3) + (x_1 + x_2 - 2x_3)H\,,$$

wobei

$$H = x_1 + x_2 - 2x_3 + \frac{1}{2}y_1^2 - 2x_3^2 + 2x_1x_2$$

gesetzt ist. Da aufgrund von (42) die Abschätzung

$$2H = (2x_1 + 1)(2x_2 + 1) - (2x_3 + 1)^2 + y_1^2 \geqq 0$$

gilt, folgt wieder $F \geqq 0$. Damit ist Lemma 4 bewiesen und zugleich der Beweis von Satz 1 beendet.

§ 2. Notwendige Ungleichungen

Definitionen: 1. Das System der 9 Ungleichungen (7) *und* (8) *heiße Ω_0, das System der 19 Ungleichungen* (9) *und* (10) *heiße Ω_1, und das System der 28 Ungleichungen aus Ω_0 und Ω_1 heiße Ω.*

2. Jeder Ungleichung aus Ω ist eine Fläche zugeordnet: Man erhält diese Fläche, indem in der betreffenden Ungleichung speziell das Gleichheitszeichen betrachtet wird. Das System der Flächen, die den Ungleichungen aus Ω bzw. Ω_k ($k = 0, 1$) zugeordnet sind, heiße Φ bzw. Φ_k.

3. Es sei Z ein beliebiger Randpunkt von $\mathfrak{F}$. Wenn eine der Flächen aus Φ den Punkt Z enthält und in der Umgebung von Z ein fünfdimensionales Stück mit $\mathfrak{F}$ gemeinsam hat, so möge sie Randfläche von $\mathfrak{F}$ im Punkte Z heißen.

4. Ist eine Fläche aus Φ Randfläche von $\mathfrak{F}$ in mindestens einem Randpunkt von $\mathfrak{F}$, so möge sie schlechthin Randfläche von $\mathfrak{F}$ heißen.

Ist $Z = X + iY$ ein beliebiger fester Punkt aus $\mathfrak{H}$, so gehört für jedes reelle $\lambda > 0$ auch der Punkt $Z(\lambda) = X + i\lambda Y$ zu $\mathfrak{H}$. Aufgrund von (15) und (26) ist der Ausdruck

$$F(\lambda) = \operatorname{abs}(CZ(\lambda) + D)$$

für jedes teilerfremde symmetrische Paar C, D mit *Rang* $C > 0$ eine monoton wachsende Funktion von λ, die für $\lambda \to \infty$ über alle Schranken wächst. Das hat zur Folge, daß die Ungleichungen (6) im Punkte $Z(\lambda)$ für hinreichend großes λ sämtlich erfüllt sind. Mit Hilfe dieser Tatsache werden wir drei Behauptungen beweisen:

I. Die Ungleichungen aus Ω_0 sind für die Definition von $\mathfrak{F}$ notwendig.

II. Die Flächen aus Φ_0 sind Randflächen von $\mathfrak{F}$.

III. Die beiden Aussagen: „Die Ungleichungen aus Ω_1 sind für die Definition von $\mathfrak{F}$ notwendig" und „Die Flächen aus Φ_1 sind Randflächen von $\mathfrak{F}$" sind äquivalent.

Um den Beweis von Satz 2 vollständig zu machen, wird dann nur noch die Notwendigkeit der Ungleichungen aus Ω_1 zu zeigen sein. Das wird folgendermaßen geschehen: Zu jeder Ungleichung U_1 aus Ω_1 wird ein Randpunkt Z von $\mathfrak{F}$ und eine in Z beginnende Kurve angegeben werden. Diese Kurve wird die Eigenschaft haben, daß die Ungleichung U_1 in allen Kurvenpunkten, die hinreichend nahe bei Z liegen, nicht erfüllt ist, die anderen Ungleichungen aus Ω aber erfüllt sind. Die dabei verwendeten Randpunkte von $\mathfrak{F}$ sind gewisse, einfach zu bestimmende Fixpunkte der Modulgruppe. Nur für die beiden Ungleichungen (10) mit $S = \begin{pmatrix} e & 0 \\ 0 & -e \end{pmatrix}$ wird ein anderer Weg eingeschlagen werden.

Es sei U_0 eine Ungleichung aus Ω_0. Dann wähle man zunächst Z_0 so, daß in Z_0 alle Ungleichungen aus Ω_0 bis auf die eine, U_0, erfüllt sind; das ist offenbar möglich. Daraufhin wähle man λ so groß, daß für $Z_0(\lambda)$ auch noch die Ungleichungen aus Ω_1 gelten. Im Punkte $Z_0(\lambda)$ sind dann alle Ungleichungen aus Ω mit Ausnahme der einen Ungleichung U_0 erfüllt. Das bedeutet die Notwendigkeit von U_0, und die Behauptung I ist bewiesen.

Offenbar kann ferner Z auf ∞^5 Weisen so gewählt werden, daß die Ungleichung U_0 in Z mit dem Gleichheitszeichen gilt und die übrigen Ungleichungen aus Ω_0 mit dem Größerzeichen erfüllt sind. Indem zu jeder Wahl von Z der Parameter λ so groß gewählt wird, daß in $Z(\lambda)$ sämtliche Ungleichungen aus Ω_1 mit dem Größerzeichen erfüllt sind, erkennt man, daß die Flächen aus Φ_0 Randflächen von $\mathfrak{F}$ sind. Damit ist die Behauptung II bewiesen.

Zum Beweis der Behauptung III setzen wir zuerst voraus, eine Ungleichung U_1 aus Ω_1 sei notwendig für die Definition von $\mathfrak{F}$. Dann existiert ein Punkt Z_1 in $\mathfrak{H}$, in welchem U_1 nicht erfüllt ist, die übrigen Ungleichungen aus Ω aber erfüllt sind. Aus Stetigkeitsgründen können wir annehmen, daß die übrigen Ungleichungen aus Ω sogar mit dem Größerzeichen gelten. Dann ist U_1 sogar in einer vollen Umgebung von Z_1 nicht erfüllt, während die übrigen Ungleichungen aus Ω dort erfüllt sind. Ist nun Z ein beliebiger Punkt aus dieser Umgebung von Z_1, so bilden wir $Z(\lambda)$ und wählen $\lambda > 1$ gerade so, daß U_1 in $Z(\lambda)$ mit dem Gleichheitszeichen gilt. Die übrigen Ungleichungen aus Ω sind in $Z(\lambda)$ mit dem Größerzeichen erfüllt. Die so gefundenen Punkte $Z(\lambda)$ bilden eine fünfdimensionale Punktmenge, sie liegen sämtlich auf der Fläche F_1, die der Ungleichung U_1 entspricht, und sie gehören zu $\mathfrak{F}$. Damit ist gezeigt, daß F_1 Randfläche von $\mathfrak{F}$ ist. Setzen wir umgekehrt voraus, daß F_1 Randfläche von $\mathfrak{F}$ ist, so gibt es einen Punkt, in dem U_1 mit dem Gleichheitszeichen erfüllt ist, die übrigen Ungleichungen aus Ω aber mit dem Größerzeichen. Aus Stetigkeitsgründen folgt die Notwendigkeit von U_1 für die Definition von $\mathfrak{F}$. Damit ist auch die Behauptung III bewiesen.

Lemma 5: Die beiden Flächen $|z_1| = 1$ *und* $\operatorname{abs} Z = 1$ *sind Randflächen von* $\mathfrak{F}$ *im Punkte* $Z_1 = iE$.

Beweis: Der Punkt Z_1 erfüllt alle Ungleichungen aus Ω. Das Gleichheitszeichen steht genau in den folgenden Fällen:

$$(44)\qquad y_3 \geq 0, \quad y_2 \geq y_1, \quad |z_1| \geq 1, \quad |z_2| \geq 1, \quad \operatorname{abs} Z \geq 1 .$$

Ist $\varepsilon > 0$ hinreichend klein, so gilt im Punkte

$$Z(\varepsilon) = i\begin{pmatrix} 1-\varepsilon & 0 \\ 0 & (1-\varepsilon)^{-1} \end{pmatrix}$$

die Ungleichung $|z_1| < 1$; die übrigen der Ungleichungen (44) bleiben erfüllt. Ebenso bleiben die Ungleichungen aus Ω, die in (44) nicht vorkommen, erfüllt, weil sie in Z_1 mit dem Größerzeichen gelten und

$$\lim_{\varepsilon \to 0} Z(\varepsilon) = Z_1$$

ist. Dieser letzte Schluß wird in allen folgenden Beweisen wiederkehren und soll daher nicht mehr besonders erwähnt werden. Damit ist die Behauptung für die Fläche $|z_1| = 1$ bewiesen. Für die Fläche $\operatorname{abs} Z = 1$ ergibt sich die Behauptung, wenn man

$$Z(\varepsilon) = i\begin{pmatrix} 1 & \varepsilon \\ \varepsilon & 1 \end{pmatrix}$$

definiert.

Lemma 6: Die Fläche $|z_2| = 1$ *ist Randfläche von* $\mathfrak{F}$ *im Punkte*

$$Z_2 = \begin{pmatrix} \varrho & 0 \\ 0 & i \end{pmatrix} \qquad \left(\varrho = -\frac{1}{2} + \frac{i}{2}\sqrt{3}\right).$$

Beweis: Der Punkt Z_2 befriedigt alle Ungleichungen aus Ω; das Gleichheitszeichen steht genau in

$$(45)\qquad x_1 \geq -\frac{1}{2}, \quad y_3 \geq 0, \quad |z_1| \geq 1, \quad |z_2| \geq 1,$$
$$\operatorname{abs}(Z + S) \geq 1 \quad \textit{für} \quad S = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

Für hinreichend kleines $\varepsilon > 0$ gilt im Punkte

$$Z(\varepsilon) = \begin{pmatrix} -\frac{1}{2} & 0 \\ 0 & 0 \end{pmatrix} + i\begin{pmatrix} \frac{1}{2}\sqrt{3} + 2\varepsilon & 0 \\ 0 & 1-\varepsilon \end{pmatrix}$$

die Ungleichung $|z_2| < 1$, während alle anderen der Ungleichungen (45) erfüllt bleiben. Insbesondere gilt für die beiden letzten dieser Ungleichungen, wenn nur das absolute Glied und lineare Glieder in ε berücksichtigt werden:

$$\operatorname{abs}(Z(\varepsilon) + S)^2 = 1 + 2\varepsilon(\sqrt{3} - 1) + O(\varepsilon^2) > 1 .$$

Lemma 7: Die Fläche $|z_1 + z_2 - 2z_3 + e| = 1$ $(e = \pm 1)$ *ist Randfläche von* $\mathfrak{F}$ *im Punkte*

$$Z_3 = \frac{e}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} + \frac{i}{2}\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}.$$

Beweis: Der Beweis wird für den Fall $e = 1$ durchgeführt. In diesem Falle erfüllt Z_3 alle Ungleichungen aus Ω, und zwar mit dem Gleichheitszeichen

genau die folgenden:

$$\frac{1}{2} \geqq x_3, \quad y_1 \geqq 2y_3, \quad y_2 \geqq y_1, \tag{46}$$
$$|z_1| \geqq 1, \quad |z_2| \geqq 1, \quad |z_1 + z_2 - 2z_3 + 1| \geqq 1.$$

Für hinreichend kleines $\varepsilon > 0$ gelten im Punkte

$$Z(\varepsilon) = \frac{1}{2}\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix} + \frac{i}{2}(1-\varepsilon)\begin{pmatrix}2 & 1\\ 1 & 2\end{pmatrix}$$

die Ungleichungen

$$|z_1| < 1, \quad |z_2| < 1, \quad |z_1 + z_2 - 2z_3 + 1| < 1,$$

während die anderen der Ungleichungen (46) erhalten bleiben. Man erreicht wieder $|z_1| \geqq 1$ und $|z_2| \geqq 1$, indem man x_1 und x_2 von Null an mit entgegengesetztem Vorzeichen wachsen läßt, derart, daß stets $x_1 + x_2 = 0$ bleibt. Die Ungleichung $|z_1 + z_2 - 2z_3 + 1| < 1$ ändert sich dabei nicht.

Die beiden folgenden Lemmas sollen gemeinsam bewiesen werden:

Lemma 8: Die beiden Flächen $\mathrm{abs}(Z + S) = 1$ *mit*

$$S = \begin{pmatrix}0 & 0\\ 0 & 0\end{pmatrix}, \begin{pmatrix}e & 0\\ 0 & e\end{pmatrix} \qquad (e = \pm 1)$$

sind Randflächen von $\mathfrak{F}$ *im Punkte* $Z_4 = e\varrho^e E$.

Lemma 9: Die beiden Flächen $\mathrm{abs}(Z + S) = 1$ *mit*

$$S = \begin{pmatrix}e & 0\\ 0 & 0\end{pmatrix}, \begin{pmatrix}0 & 0\\ 0 & -e\end{pmatrix} \qquad (e = \pm 1)$$

sind Randflächen von $\mathfrak{F}$ *im Punkte*

$$Z_5 = \begin{pmatrix}e\varrho^e & 0\\ 0 & -e\varrho^{-e}\end{pmatrix}.$$

Beweis: Der Beweis wird für den Fall $e = 1$ durchgeführt. Die beiden Punkte $Z_4 = \varrho E$ und $Z_5 = \begin{pmatrix}\varrho & 0\\ 0 & -\varrho^{-1}\end{pmatrix}$ genügen allen Ungleichungen aus Ω. Das Gleichheitszeichen steht für beide Punkte genau in folgenden Ungleichungen:

$$x_1 \geqq -\frac{1}{2}, \quad y_3 \geqq 0, \quad y_2 \geqq y_1, \quad |z_1| \geqq 1, \quad |z_2| \geqq 1. \tag{47}$$

Für Z_4 steht das Gleichheitszeichen außerdem noch in

$$x_2 \geqq -\frac{1}{2}, \quad \mathrm{abs}(Z + S) \geqq 1 \quad \textit{für} \quad S = \begin{pmatrix}0 & 0\\ 0 & 0\end{pmatrix}, \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}, \begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}, \begin{pmatrix}0 & 0\\ 0 & 1\end{pmatrix}; \tag{48}$$

und für Z_5 steht das Gleichheitszeichen außer in (47) noch in

$$\frac{1}{2} \geqq x_2, \quad \mathrm{abs}(Z + S) \geqq 1 \quad \textit{für} \quad S = \begin{pmatrix}0 & 0\\ 0 & -1\end{pmatrix}, \begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}, \begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}, \begin{pmatrix}0 & 0\\ 0 & 0\end{pmatrix}. \tag{49}$$

Es sei nun ε eine positive oder negative Zahl, deren Absolutbetrag hinreichend klein ist. Für $Z^* = Z_4, Z_5$ bilden wir

$$Z(\varepsilon) = Z^* + (\varepsilon + i|\varepsilon|)\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix} = \begin{pmatrix}-\frac{1}{2} & \varepsilon\\ \varepsilon & \mp\frac{1}{2}\end{pmatrix} + i\begin{pmatrix}\frac{1}{2}\sqrt{3} & |\varepsilon|\\ |\varepsilon| & \frac{1}{2}\sqrt{3}\end{pmatrix}.$$

Für diesen Punkt bleiben alle Ungleichungen (47) und jeweils die erste der Ungleichungen (48) und (49) erhalten. Bezeichnen wir ferner den Realteil des Punktes $Z(\varepsilon) + S$ mit $R = \begin{pmatrix} r_1 & \varepsilon \\ \varepsilon & r_2 \end{pmatrix}$, so wird

$$(50) \qquad \operatorname{abs}(Z(\varepsilon) + S)^2 = \left(\frac{3}{4} - r_1 r_2\right)^2 + \left(\frac{1}{2}\sqrt{3}\,(r_1 + r_2) - 2\varepsilon|\varepsilon|\right)^2 .$$

Für $Z^* = Z_4$ ist dieser Ausdruck aufgrund von (48) in den folgenden vier Fällen

$$(51) \qquad R_1 = \begin{pmatrix} -\frac{1}{2} & \varepsilon \\ \varepsilon & -\frac{1}{2} \end{pmatrix}, \quad R_2 = \begin{pmatrix} \frac{1}{2} & \varepsilon \\ \varepsilon & \frac{1}{2} \end{pmatrix}, \quad R_3 = \begin{pmatrix} \frac{1}{2} & \varepsilon \\ \varepsilon & -\frac{1}{2} \end{pmatrix}, \quad R_4 = \begin{pmatrix} -\frac{1}{2} & \varepsilon \\ \varepsilon & \frac{1}{2} \end{pmatrix}$$

zu untersuchen. Für $\varepsilon < 0$ wird

$$(52) \qquad \operatorname{abs}(R_1 + i\,Y(\varepsilon)) < 1 , \quad \operatorname{abs}(R_k + i\,Y(\varepsilon)) > 1 \qquad (k = 2,3,4).$$

Das bedeutet, daß $\operatorname{abs} Z = 1$ Randfläche von $\mathfrak{F}$ im Punkte Z_4 ist. Für $\varepsilon > 0$ wird

$$(53) \qquad \operatorname{abs}(R_2 + i\,Y(\varepsilon)) < 1 , \quad \operatorname{abs}(R_k + i\,Y(\varepsilon)) > 1 \qquad (k = 1,3,4).$$

Das bedeutet, daß die Fläche $\operatorname{abs}(Z + S) = 1$ mit $S = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ Randfläche von $\mathfrak{F}$ im Punkte Z_4 ist. Damit ist Lemma 8 bewiesen. Für $Z^* = Z_5$ ist der Ausdruck (50) aufgrund von (49) ebenfalls in den vier Fällen (51) zu untersuchen. Jetzt bedeutet aber (52), daß die Fläche $\operatorname{abs}(Z + S) = 1$ mit $S = \begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix}$ Randfläche von $\mathfrak{F}$ im Punkte Z_5 ist; und (53) bedeutet dasselbe für die Fläche $\operatorname{abs}(Z + S) = 1$ mit $S = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$. Damit ist auch Lemma 9 bewiesen.

Lemma 10: Die Fläche $\operatorname{abs}(Z + S) = 1$ *mit* $S = \begin{pmatrix} e & e \\ e & 0 \end{pmatrix}$ $(e = \pm 1)$ *ist Randfläche von* $\mathfrak{F}$ *im Punkte*

$$Z_6 = -\frac{e}{2}\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} + \frac{i}{2}\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}.$$

Der Beweis dieses Lemmas verläuft völlig analog zum Beweis des nächsten, es müssen nur die beiden Variablen z_1 und z_2 vertauscht werden, so daß nur dieses folgende Lemma bewiesen werden soll.

Lemma 11: Die Fläche $\operatorname{abs}(Z + S) = 1$ *mit* $S = \begin{pmatrix} 0 & e \\ e & e \end{pmatrix}$ $(e = \pm 1)$ *ist Randfläche von* $\mathfrak{F}$ *im Punkte*

$$Z_7 = -\frac{e}{2}\begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} + \frac{i}{2}\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}.$$

Beweis: Der Beweis wird für $e = 1$ durchgeführt. Der Punkt Z_7 erfüllt dann alle Ungleichungen aus Ω, und zwar mit dem Gleichheitszeichen genau die folgenden:

$$(54) \qquad x_2 \geqq -\frac{1}{2}, \quad x_3 \geqq -\frac{1}{2}, \quad y_1 \geqq 2y_3, \quad y_2 \geqq y_1, \quad |z_1| \geqq 1$$

und

$$(55) \qquad \operatorname{abs}(Z + S) \geqq 1 \quad \textit{mit} \quad S = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \ \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}.$$

Für hinreichend kleines $\varepsilon > 0$ wird der Punkt

$$Z(\varepsilon) = \frac{1}{2}\begin{pmatrix} 2\varepsilon & -1 \\ -1 & -1 \end{pmatrix} + \frac{i}{2}\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$$

betrachtet. Für ihn bleiben die Ungleichungen (54) erhalten. Für die Ungleichungen (55) folgt im Falle $S = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$ die Abschätzung

$$\operatorname{abs} Z^2(\varepsilon) = \left(1 + \frac{1}{2}\varepsilon\right)^2 + \varepsilon^2 > 1$$

und im Falle $S = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$ die Abschätzung

$$\operatorname{abs}(Z(\varepsilon) + S)^2 = \left(1 - \frac{1}{2}\varepsilon\right)^2 + \varepsilon^2 = 1 - \varepsilon + \frac{5}{4}\varepsilon^2 < 1 .$$

Lemma 12: Die Fläche $\operatorname{abs}(Z + S) = 1$ *mit* $S = \begin{pmatrix} 0 & e \\ e & 0 \end{pmatrix}$ $(e = \pm 1)$ *ist Randfläche von* $\mathfrak{F}$ *im Punkte*

$$Z_8 = \frac{e}{3}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} + \frac{i}{3}\sqrt{2}\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}.$$

Beweis: Der Beweis wird für $e = 1$ durchgeführt. Der Punkt Z_8 erfüllt dann alle Ungleichungen aus Ω, und zwar mit dem Gleichheitszeichen genau die folgenden:

(56) $\qquad y_1 \geqq 2y_3, \quad y_2 \geqq y_1, \quad |z_1| \geqq 1, \quad |z_2| \geqq 1, \quad |z_1 + z_2 - 2z_3 - 1| \geqq 1$

und

(57) $\qquad \operatorname{abs}(Z + S) \geqq 1$ *für* $S = \begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix}, \begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$

Sei $\varepsilon > 0$ hinreichend klein und a eine positive noch zu bestimmende Zahl. Für den Punkt

$$Z(\varepsilon) = \frac{1}{3}\begin{pmatrix} 1 & -1 - 3\varepsilon \\ -1 - 3\varepsilon & 1 \end{pmatrix} + \frac{i}{3}\sqrt{2}\begin{pmatrix} 2 & 1 - 2a\varepsilon \\ 1 - 2a\varepsilon & 2 \end{pmatrix}$$

bleiben dann die Ungleichungen (56) erhalten. Bezeichnen wir ferner den Realteil von $Z(\varepsilon) + S$ mit

$$R - \varepsilon\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} r_1 & r_3 \\ r_3 & r_2 \end{pmatrix} - \varepsilon\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

so wird, wenn ε^2 und noch höhere Potenzen von ε vernachlässigt werden:

$$F = \operatorname{abs}(Z(\varepsilon) + S)^2 = \left(\frac{2}{3} + r_3^2 - r_1 r_2 + \varepsilon\left(\frac{8}{9}a - 2r_3\right)\right)^2 +$$
$$+ \frac{8}{9}(r_1 + r_2 - r_3 + \varepsilon(1 + 2ar_3))^2 + O(\varepsilon^2) .$$

Dieser Ausdruck ist nach (57) in den folgenden vier Fällen

$$R_1 = \frac{1}{3}\begin{pmatrix} 1 & -1 \\ -1 & -2 \end{pmatrix}, \quad R_2 = \frac{1}{3}\begin{pmatrix} -2 & -1 \\ -1 & 1 \end{pmatrix}, \quad R_3 = \frac{1}{3}\begin{pmatrix} -2 & -1 \\ -1 & -2 \end{pmatrix}, \quad R_4 = \frac{1}{3}\begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$$

zu untersuchen. Für R_k mit $k = 1, 2$ wird

$$F_k = 1 + 2\varepsilon\left(\frac{8}{9}a + \frac{2}{3}\right) + O(\varepsilon^2) > 1 .$$

Für R_3 ergibt sich die Gleichung

$$F_3 = 1 + 2\varepsilon\left(\frac{8}{9}a - \frac{2}{3}\right) + O(\varepsilon^2),$$

und schließlich für R_4 die Gleichung

$$F_4 = 1 + 2\varepsilon\left(\frac{8}{9}a - \frac{4}{3}\right) + O(\varepsilon^2).$$

Für alle Zahlen a mit $\frac{3}{4} < a < \frac{3}{2}$ und hinreichend kleines $\varepsilon > 0$ gilt also $F_3 > 1$ und $F_4 < 1$, womit Lemma 12 bewiesen ist.

Wir müssen jetzt noch zeigen, daß auch die beiden Ungleichungen

$$\operatorname{abs}(Z + S) \geqq 1 \qquad S = \begin{pmatrix} e & 0 \\ 0 & -e \end{pmatrix},\ e = \pm 1 \tag{58}$$

zur Definition von $\mathfrak{F}$ notwendig sind. Aufgrund von Lemma 1 genügt es, den Fall $e = 1$ zu behandeln. Wir betrachten die Punkte

$$Z = \begin{pmatrix} -\frac{1}{2} & x \\ x & \frac{1}{2} \end{pmatrix} + i\begin{pmatrix} \frac{1}{2}\sqrt{3} & \frac{1}{4}\sqrt{3} \\ \frac{1}{4}\sqrt{3} & y \end{pmatrix} \qquad \left(0 \leqq x \leqq \frac{1}{8},\ y \geqq 1\right). \tag{59}$$

Offenbar erfüllen diese Punkte alle Ungleichungen aus Ω_0. Abgesehen von den drei Ungleichungen

$$\operatorname{abs}(Z + S_k) \geqq 1 \qquad (k = 1,2,3) \tag{60}$$

mit

$$S_1 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix},\quad S_2 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix},\quad S_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

genügen die Punkte (59) aber auch den Ungleichungen aus Ω_1, wie jetzt gezeigt werden soll. Die vier Ungleichungen (9) sind erfüllt. Mit $S = \begin{pmatrix} s_1 & s_3 \\ s_3 & s_2 \end{pmatrix}$ wird

$$\operatorname{abs}(Z + S)^2 = \left(-\left(s_1 - \frac{1}{2}\right)\left(s_2 + \frac{1}{2}\right) + (s_3 + x)^2 + \frac{1}{2}\, y\sqrt{3} - \frac{3}{16}\right)^2 + \tag{61}$$
$$+ \left(y\left(s_1 - \frac{1}{2}\right) + \frac{1}{2}\sqrt{3}\left(s_2 + \frac{1}{2}\right) - \frac{1}{2}\sqrt{3}\,(s_3 + x)\right)^2.$$

In den Fällen $S = \begin{pmatrix} -1 & * \\ * & 0 \end{pmatrix}, \begin{pmatrix} -1 & * \\ * & 1 \end{pmatrix}, \begin{pmatrix} 0 & * \\ * & 1 \end{pmatrix}$ beachten wir nur das erste Quadrat auf der rechten Seite:

$$\operatorname{abs}(Z + S) \geqq \frac{3}{4} + \frac{1}{2}\sqrt{3} - \frac{3}{16} > 1.$$

Auch in den Fällen $S = \begin{pmatrix} 0 & e \\ e & 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & -1 \\ -1 & -1 \end{pmatrix}$ beachten wir in (61) rechts nur das erste Quadrat:

$$\operatorname{abs}(Z + S) \geqq -\frac{1}{4} + (1 - x)^2 + \frac{1}{2}\sqrt{3} - \frac{3}{16} \geqq \frac{21}{64} + \frac{1}{2}\sqrt{3} > 1.$$

Es bleiben noch die drei Fälle $S_4 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $S_5 = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$, $S_6 = \begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix}$ zu untersuchen. In den ersten beiden dieser Fälle beachten wir in (61) rechts nur das zweite Quadrat:

$$\operatorname{abs}(Z + S_4) \geqq \frac{1}{2} + \frac{3}{4}\sqrt{3} - \frac{1}{16}\sqrt{3} > 1,$$

$$\operatorname{abs}(Z + S_5) \geqq \frac{3}{2} + \frac{1}{4}\sqrt{3} > 1.$$

Im letzten Fall wird schließlich

$$\mathrm{abs}(Z + S_6)^2 = \left(-\frac{1}{4} + x^2 + \frac{1}{2}\,y\sqrt{3} - \frac{3}{16}\right)^2 + \left(-\frac{1}{2}\,y - \frac{1}{4}\sqrt{3} - \frac{1}{2}\,x\sqrt{3}\right)^2 \geq$$

$$\geq \left(\frac{1}{2}\sqrt{3} - \frac{1}{2}\right)^2 + \left(\frac{1}{2} + \frac{1}{4}\sqrt{3}\right)^2 = 1 + \frac{1}{16}(7 - 4\sqrt{3}) > 1\,.$$

Es handelt sich jetzt darum, die drei kritischen Ungleichungen (60) zu untersuchen. Im folgenden werden die beiden positiven Zahlen

$$\xi_1 = \frac{1}{4}\sqrt{15} - \frac{1}{2}\sqrt{3}\,, \qquad \xi_2 = \frac{1}{2}\sqrt{3} - \frac{3}{4}$$

benötigt. Sie stehen in der folgenden Größenbeziehung zueinander:

$$0 < \xi_1 < \xi_2 < \frac{1}{8}\,.$$

Wir setzen

(62) $$H_k = \mathrm{abs}(Z + S_k)^2 \qquad (k = 1, 2, 3)$$

und fragen, für welche x die Beziehung $H_2 > H_3$ gilt. Es ist

$$H_2 - H_3 = -x^2 - \frac{3}{2}\,x + \frac{3}{16}\,.$$

Weil die positive Lösung der quadratischen Gleichung $x^2 + \frac{3}{2}\,x - \frac{3}{16} = 0$ die Zahl $x = \xi_2$ ist, gilt $H_2 > H_3$ sicher für $0 \leq x < \xi_2$. Wir fragen ferner danach, für welche x die Ungleichung $H_1 > H_2$ im ganzen Intervall $y \geq 1$ erfüllt ist. Es gilt

$$H_1 - H_2 = x^2 + xy\sqrt{3} - \frac{3}{16} \geq x^2 + x\sqrt{3} - \frac{3}{16}.$$

Die positive Lösung der quadratischen Gleichung $x^2 + x\sqrt{3} - \frac{3}{16} = 0$ ist $x = \xi_1$. Beschränken wir daher x auf das Intervall

(63) $$\xi_1 < x < \xi_2\,,$$

so gilt für alle $y \geq 1$ die Beziehung

(64) $$H_1 > H_2 > H_3\,.$$

Wir untersuchen jetzt den Ausdruck H_2 genauer und bilden dazu die partiellen Ableitungen nach x und y:

(65) $$H_2 = \left(x^2 + \frac{1}{2}\,y\sqrt{3} - \frac{7}{16}\right)^2 + \left(\frac{1}{2}\,y + \frac{1}{4}\sqrt{3} - \frac{1}{2}\,x\sqrt{3}\right)^2,$$

(66) $$\frac{\partial H_2}{\partial x} = 4x\left(x^2 + \frac{1}{2}\,y\sqrt{3} - \frac{7}{16}\right) - \sqrt{3}\left(\frac{1}{2}\,y + \frac{1}{4}\sqrt{3} - \frac{1}{2}\,x\sqrt{3}\right),$$

(67) $$\frac{\partial H_2}{\partial y} = \sqrt{3}\left(x^2 + \frac{1}{2}\,y\sqrt{3} - \frac{7}{16}\right) + \left(\frac{1}{2}\,y + \frac{1}{4}\sqrt{3} - \frac{1}{2}x\sqrt{3}\right).$$

Aus (66) folgt sofort

(68) $$\frac{\partial H_2}{\partial x} < 0 \qquad \textit{für} \quad y = 1,\ x = \frac{1}{8}\,.$$

Wir beweisen ferner

(69) $$H_2 < 1 \qquad \textit{für} \quad y = 1,\ x = \xi_1\,.$$

Nach (65) gilt nämlich für die Werte $y = 1, x = \frac{1}{4}\sqrt{15} - \frac{1}{2}\sqrt{3}$ die Gleichung

$$H_2 = \left(\frac{5}{4} - \frac{3}{4}\sqrt{5} + \frac{1}{2}\sqrt{3}\right)^2 + \left(\frac{5}{4} - \frac{3}{8}\sqrt{5} + \frac{1}{4}\sqrt{3}\right)^2 =$$

$$= 1 + \frac{1}{64}\left(421 + 120\sqrt{3} - 180\sqrt{5} - 60\sqrt{15}\right).$$

Beachten wir $\sqrt{5} > \frac{89}{40}$, so folgt

$$H_2 < 1 + \frac{1}{64}\left(421 + 120\sqrt{3} - \frac{9\cdot 89}{2} - \frac{3\cdot 89}{2}\sqrt{3}\right).$$

Wegen $\frac{9\cdot 89}{2} > 400$ und $\frac{3\cdot 89}{2} > 133$ ergibt sich schließlich

$$H_2 < 1 + \frac{1}{64}\left(21 - 13\sqrt{3}\right) < 1.$$

Damit ist (69) bewiesen.

Da der Ausdruck $\frac{\partial H_2}{\partial x}$ nach (66) für $y = 1$ eine monoton wachsende Funktion von x ist, die für $x = \frac{1}{8}$ nach (68) kleiner als Null ist, gilt

$$\frac{\partial H_2}{\partial x} < 0 \qquad \textit{für} \quad y = 1, 0 \leqq x \leqq \frac{1}{8}.$$

Also ist H_2 für $y = 1$ im ganzen Intervall $0 \leqq x \leqq \frac{1}{8}$ eine monoton fallende Funktion von x. Hieraus folgt unter Benutzung von (69) die Ungleichung

$$H_2 < 1 \qquad \textit{für} \quad y = 1, \xi_1 < x < \xi_2. \tag{70}$$

Nach (67) ist für $y \geqq 1$ und $\xi_1 < x < \xi_2$ der Ausdruck $\frac{\partial H_2}{\partial y} > 0$. Also ist H_2 für jedes x aus dem Intervall (63) eine monoton wachsende Funktion von y. Für $y = 1$ ist diese nach (70) kleiner als 1. Wir vergrößern dann y gerade soweit, bis $H_2 = 1$ wird. Nach (64) gilt dann für jedes x aus dem Intervall (63) die Beziehung

$$H_1 > H_2 = 1 > H_3.$$

Beachten wir jetzt (62), so folgt die Existenz eines Punktes (59), der alle Ungleichungen aus Ω mit Ausnahme der einen Ungleichung (60) für $k = 3$ erfüllt. Das besagt gerade die Notwendigkeit der Ungleichung (58) für $e = 1$.

Damit ist die Notwendigkeit der Ungleichungen aus Ω gezeigt und der Beweis von Satz 2 beendet.

Literatur

[1] Siegel, C. L.: Einführung in die Theorie der Modulfunktionen n-ten Grades. Math. Ann. **116**, 617—657 (1939). — [2] Siegel, C. L.: Symplectic Geometry. Amer. J. Math. **65**, 1—86 (1943). — [3] Minkowski, H.: Diskontinuitätsbereich für arithmetische Äquivalenz. J. reine angew. Math. **129**, 220—274 (1905). — [4] Siegel, C. L.: Einheiten quadratischer Formen. Abh. Math. Sem. Hansischen Univ. **13**, 209—239 (1940). — [5] Dedekind, R.: Schreiben an Herrn Borchardt über die Theorie der elliptischen Modulfunktionen. J. reine angew. Math. **83**, 265—292 (1877).

(Eingegangen am 5. Februar 1959)

Rogers, H. jr.
Math. Annalen 138, 125—140 (1959)

Computing Degrees of Unsolvability

By

Hartley Rogers jr. in Cambridge (Mass.)*

Introduction

In [1] and [2] the arithmetical sets are classified in the Kleene-Mostowski hierarchy of predicate forms. In [3] Post defines certain reducibility notions; these give rise to the degrees of unsolvability discussed in [4] by Kleene and Post. In [5] Turing shows the unsolvability of the halting problem for Turing machines; in this case and in the case of other early examples of unsolvability, the unsolvable problems presented concern the determination of certain facts about a partial recursive function, given defining equations for that function. The present paper has to do with problems of this kind and with their classification as to degree of unsolvability. Notions from the theory of the predicate hierarchy and from the theory of reducibility are used to locate and measure the problems considered. The paper falls into two parts.

Part I is, in part, expository. Basic facts about degrees of unsolvability are summarized, and the relation of the degree of a set to its position in the predicate hierarchy is discussed.

Let φ_x be the partial recursive function whose defining equations have Gödel number x, $x = 0, 1, 2, \ldots$. Let W_x be the range of φ_x. In Part II, we consider the problems represented by the sets $\{x|W_x$ is recursive$\}$, $\{x|W_x$ is simple$\}$, $\{x|W_x$ is hypersimple$\}$, $\{x|\overline{W}_x$ is finite$\}$ and $\{x|W_x$ is creative$\}$; and we calculate the degree of unsolvability for each set. In fact, we locate each set up to recursive isomorphism type. A partial result is obtained for $\{x|W_x$ is complete$\}$, and the conjecture of Davis and Shapiro is verified that $\{x|W_x$ is complete$\}$ and $\{x|W_x$ is not recursive$\}$ have different degree of unsolvability (with respect to many-one reducibility). The solution to Post's Problem is a corollary to this result. (The notions *simple, complete,* etc. were defined in [3].) The constructions used owe much to the methods of Friedberg in [6]. The method of computation can be summarized as follows. The Kleene-Mostowski predicate hierarchy, together with methods of Tarski and Kuratowski, is used to get upper bounds on degree of unsolvability. With certain distinguished degrees are associated certain *standard* problems of relatively simple intrinsic significance. Reducibility from these standard problems is then used to get lower bounds on degree.

The author is indebted to Norman Shapiro, whose comments initiated his interest in this area; to Andrzej Mostowski, who, in paragraph 3, page 260

*) Mass. Institute of Technology.

of [7], provided stimulus to further exploration; and to JOHN ADDISON who suggested the immediate problems treated.

Part I

§ 1. The predicate hierarchy

Notation. Let N be the set of non-negative integers. 'A', 'B', ... shall denote subsets of N, and 'x', 'y', ... shall denote members of N. N^n is the cartesian product (of n factors) $N \times N \times \cdots \times N$. '$R$', '$S$', ... shall denote subsets of N^n. Henceforth the words 'integer' and 'number' refer to members of N. We define $\overline{A} =_{df} N - A$.

Elementary number theory is the formalism obtained by the use of a lower predicate calculus with equality that employs two ternary predicate constants for addition and multiplication. Taking the integers as domain and ordinary addition and multiplication as the model for this formalism, we say that a relation $R \subset N^n$ is *arithmetical* if it is explicitly definable by a formula of elementary number theory.

Let R be a subset of N^n. The relation R is said to be *recursive* if the function $f: N^n \to N$ such that $f = 1$ on R and $f = 0$ on $N^n - R$ is a recursive function. As is known from work of GÖDEL (for a recent proof see ROBINSON [8]), every recursive relation is arithmetical. Conversely, by going to prenex form, it is immediate that every arithmetical relation S can be expressed as

$$\{\langle x_1, \ldots, x_n\rangle | (Q_1 y_1) \ldots (Q_m y_m) [\langle y_1, \ldots, y_m, x_1, \ldots, x_n\rangle \in R]\}$$

where $Q_1, \ldots, Q_m$ are quantifiers and $R \subset N^{n+m}$ is a recursive relation. Using the abbreviation '$R(y_1, \ldots, x_n)$' for '$\langle y_1, \ldots, x_n\rangle \in R$', this gives the following

Lemma. *For any set A, A is arithmetical if and only if there is a recursive relation $R(\subset N^{n+1}$ for some $n \geqq 0)$ such that $A = \{x | (Q_1 y_1) \ldots (Q_n y_n) R(y_1, \ldots, y_n, x)\}$ for some (possibly empty) choice of quantifiers $Q_1, \ldots, Q_n$.*

Given $Q_1, \ldots, Q_n$, and recursive R, we speak of the assertion $(Q_1 y_1) \ldots (Q_n y_n) R(y_1, \ldots, y_n, x)$ as a *predicate form* and say that it *expresses* the set A.

Let p_i be the $i + 1^{st}$ prime. KLEENE defines $(x)_i$ to be the exponent of p_i in the prime factorization of x (with the convention that $(0)_i = 0$). Clearly the assertion $(\forall x)$ $(\forall y)$ $[\ldots x \ldots y \ldots]$ is equivalent to the assertion $(\forall x)$ $[\ldots (x)_0 \ldots (x)_1 \ldots]$, and the assertion $(\exists x)$ $(\exists y)$ $[\ldots x \ldots y \ldots]$ is equivalent to the assertion $(\exists x)$ $[\ldots (x)_0 \ldots (x)_1 \ldots]$. Furthermore if

$$\{\langle \ldots x \ldots y \ldots\rangle | R(\ldots x \ldots y \ldots)\} \subset N^n$$

is recursive, then

$$\{\langle \ldots x \ldots\rangle | R(\ldots (x)_0 \ldots (x)_1 \ldots)\} \subset N^{n-1}$$

is recursive. This allows the simplification of predicate forms by 'collapse' of adjacent quantifiers of like kind. Thus we have

Corollary. *A is arithmetical if and only if A can be expressed by a predicate form with quantifiers that alternate in kind; furthermore there is such an alternating quantifier form where the number of quantifiers is as small as the number of quantifiers in any form expressing A.*

Definition. $\Sigma_n =_{df}$ the collection of all sets expressible by a predicate form with n alternating quantifiers of which the first is existential.

$\Pi_n =_{df}$ the collection of all sets expressible by a predicate form with n alternating quantifiers of which the first is universal.

In particular, $\Sigma_0 = \Pi_0 =$ the collection of all recursive sets.

The following basic facts can be proved.

(1) $A \in \Sigma_n \Leftrightarrow \bar{A} \in \Pi_n$. This is immediate.

(2) $\Sigma_n \subset \Sigma_{n+1}, \Pi_n \subset \Sigma_{n+1}, \Sigma_n \subset \Pi_{n+1}$ and $\Pi_n \subset \Pi_{n+1}$. This is immediate by introduction of superfluous quantifiers. (E.g. the assertion $(\forall x) [\ldots x \ldots]$ is equivalent to the assertion $(\forall x) (\exists y) [[\ldots x \ldots] \; \& \; y = y]$.) The reader will note that, as a consequence of (2) and the Corollary, the word 'alternating' can be deleted from the definition for Σ_n and Π_n.

(3) $\Sigma_n \not\subset \Pi_n$ and $\Pi_n \not\subset \Sigma_n$ for $n > 0$. Thus the containments of (2) are all proper. This is the *hierarchy theorem* of Kleene [1].

(4) $\Sigma_0 = \Pi_0 = \Sigma_1 \cap \Pi_1$. It is easily shown that Σ_1 is the collection of all recursively enumerable sets, and (4) follows directly. (4) is known as Post's Theorem.

(5) By (2), $\Sigma_n \cup \Pi_n \subset \Sigma_{n+1} \cap \Pi_{n+1}$; for $n > 0$, this containment is proper. (Take $A \in \Sigma_n - \Pi_n$. Take $B =_{df} \{x | (\exists y) [[x = 2y \; \& \; y \in A] \vee [x = 2y + 1 \; \& \; y \in \bar{A}]]\}$. As we shall see below, $B \in \Sigma_{n+1} \cap \Pi_{n+1}$, but $B \notin \Sigma_n \cup \Pi_n$.)

The Σ and Π classes are sometimes said to give a classification of the arithmetical sets according to the *hierarchy of predicate forms*.

§ 2. Degrees of unsolvability

Notation. 'f', 'g', ... denote functions defined on N and mapping into N. 'φ', 'θ', ... denote functions defined on any subset of N and mapping into N ("partial functions"). The *partial recursive functions* are those partial functions which are computable from finite sets of defining equations within the Kleene recursive function formalism (see [1]); or, equivalently, they are those partial functions which are computable by Turing machine (see [9]). The *recursive functions* are the partial recursive functions which are defined on N. An indexing ("Gödel numbering") for the partial recursive functions can be obtained from an effective enumeration of all possible finite sets of defining equations. Let φ_x be the partial function given by the x^{th} set of equations. A different indexing can be obtained from an effective enumeration of all Turing machines. The results obtained below hold for either method of indexing or for any similar method. (For a general discussion of equivalent indexings, see [10]). W_x is the range of φ_x, i.e. the recursively enumerable set with index x.

A partial function is *a partial recursive function in* X for a given set X, if it is computable by a Turing machine which has available a black box ('oracle') that will supply, as often as desired and for any z, correct information as to whether or not $z \in X$. From an enumeration of these machines, an indexing can be obtained for the partial recursive functions in X. Let 'φ_x^X' denote the partial recursive function in X with index x, and let 'W_x^X' denote its range. This indexing is uniform in the sense that, as X varies, φ_x^X is computed

by the same machine; only the oracle changes. The Kleene formalism can also be used to give an equivalent definition and an alternative indexing. Note that for any recursive set X, $\{\varphi_z^X\}_{z=0}^{z=\infty}$ gives an indexing of the partial recursive functions.

Remark. The notion of oracle can be eliminated in the following way. If $x = 2^{x_1} + 2^{x_2} + \cdots + 2^{x_n}$, $x_1 > x_2 > \cdots > x_n$, we write D_x for $\{x_1, x_2, \ldots, x_n\}$ and call x a *canonical index* (after RICE) for the finite set D_x. We set $D_0 = \theta$, where θ is the empty set. We say that z *succeeds* for A if

$$(\exists v)\,[v \in W_z \,\&\, v = 2^s 3^t \,\&\, D_s \subset A \,\&\, D_t \subset \overline{A}]\,.$$

Define V_z^X to be $\{z | (\exists y)\,[2^y 3^z \in W_x \,\&\, y \text{ succeeds for } X]\}$. Define ψ_z^X by $\psi_z^X(z) \cong \mu y[2^y 3^z \in V_z^X]$. It is not difficult to show (see [11]) that the $\{V_z^X\}$ are the $\{W_z^X\}$ under an alternative equivalent indexing and that the $\{\psi_z^X\}$ are the $\{\varphi_z^X\}$ under an equivalent indexing.

Definitions. $A \leqq_T B =_{\mathrm{df}}$ there exist an x and an f such that

$$[f = \varphi_x^B \,\&\, f(A) = 1 \,\&\, f(\overline{A}) = 0]\,.$$

A recursively enumerable in $B =_{\mathrm{df}}$ there is an x such that $A = W_x^B$.

It is immediate that $[A \leqq_T B \Leftrightarrow A$ and $\overline{A}$ both recursively enumerable in $B]$, and it is further easily shown that the relation $\leqq_T$ is reflexive and transitive.

$A \equiv_T B =_{\mathrm{df}} A \leqq_T B$ and $B \leqq_T A$.

This is an equivalence relation on subsets of N. The relation $\leqq_T$ is called *Turing reducibility* and the equivalence classes are called *degrees of unsolvability with respect to Turing reducibility, ("T-degrees")*. The relation $\leqq_T$ induces a partial ordering on the T-degrees.

$A \leqq_1 B =_{\mathrm{df}}$ there exists a one-one recursive function f such that $f(A) \subset B$ and $f(\overline{A}) \subset \overline{B}$.

$\leqq_1$ is reflexive and transitive.

$$A \equiv_1 B =_{\mathrm{df}} A \leqq_1 B \quad\text{and}\quad B \leqq_1 A\,.$$

$\leqq_1$ is called *one-one reducibility* and induces a partial ordering on the equivalence classes of $\equiv_1$; the latter are the *degrees of unsolvability with respect to one-one reducibility ("1-degrees")*.

$\mathcal{G} =_{\mathrm{df}} \{f | f$ is a one-one recursive function mapping onto $N\}$.

$\mathcal{G}$ is a group under composition of functions, the *group of recursive permutations*, and it generates the following relation, called *recursive isomorphism*.

$$A \equiv B =_{\mathrm{df}} f(A) = B \text{ for some } f \in \mathcal{G}\,.$$

MYHILL has shown [12] that $A \equiv_1 B \Leftrightarrow A \equiv B$, i.e. 1-degrees and recursive isomorphism types coincide.

We thus have two orderings of degrees, and one is a refinement of the other in the sense that $A \leqq_1 B \Rightarrow A \leqq_T B$.

The *jump operation* is useful in investigating these two structures. It is defined by

$$A' =_{df} \{x \mid x \in W_x^A\}.$$

As we shall note, it is well-defined with respect to both T-degrees and 1-degrees.

The following basic facts can be proved about the above concepts.

(1) $A \leq_1 A'$ but not $A' \leq_T A$.

(2) A recursively enumerable in $B \Leftrightarrow A \leq_1 B'$.

(3) $A \leq_T B \Leftrightarrow A' \leq_1 B'$. This shows that the jump operation is well-defined on degrees. (1), (2) and the $\Rightarrow$ part of (3) are shown in [4]. $\Leftarrow$ for (3) uses (2) and the fact that A and $\bar{A}$ are both recursively enumerable in A; for then both A and $\bar{A}$ are $\leq_1 A'$, hence both are $\leq_1 B'$, and hence both are recursively enumerable in B.

(4) $A \equiv_T B \Leftrightarrow A' \equiv B'$. This is immediate from (3).

Definition. Let $S^{(0)} =_{df} \theta$; $S^{(n+1)} =_{df} (S^{(n)})'$. Then, by (1), $S^{(0)}$, $S^{(1)}, \ldots$ gives a sequence of successively higher T-degrees and hence, also by (1), of successively higher 1-degrees. We use '$S^{(n)}$' to denote the 1-degree of the set $S^{(n)}$. (This identification of notation is harmless since all our concepts of Σ_n, Π_n, degrees, jump, ... are recursively invariant; i.e. invariant with respect to $\mathscr{G}$.) We take $O^{(n)} =_{df}$ the T-degree of $S^{(n)}$.

Further facts about the partial orderings of degrees and about the jump can be found in [4], [6] and [13].

§ 3. Degrees and predicates

There are certain well known relationships between the degree orderings and the predicate hierarchy; see for instance [14]. We summarize them here.

(1) A recursively enumerable in $S^{(n)} \Leftrightarrow A \in \Sigma_{n+1} \Leftrightarrow A \leq_1 S^{(n+1)}$. Hence $S^{(n+1)}$ is a maximum 1-degree, under $\leq_1$, in Σ_{n+1}. Taking compliments, we see that $\overline{S^{(n+1)}}$ is maximum in Π_{n+1}. Note that $S^{(n+1)}$ is in Σ_{n+1} but not in Π_{n+1}, and $\overline{S^{(n+1)}}$ is in Π_{n+1} but not in Σ_{n+1}. (1) can be proved straightforwardly by induction on n.

(2) $A \in \Sigma_{n+1} \cap \Pi_{n+1} \Leftrightarrow A \leq_T S^{(n)}$. This follows from (1) and § 2. Hence $O^{(n)}$ is a maximum T-degree in $\Sigma_{n+1} \cap \Pi_{n+1}$. Hence, from (1) and (2), we have that the union of all 1-degrees $\leq_1 S^{(n+1)}$ is exactly Σ_{n+1} and the union of all T-degrees $\leq_T S^{(n+1)}$ is exactly $\Sigma_{n+2} \cap \Pi_{n+2}$.

(3) A is arithmetical $\Leftrightarrow A \in \Sigma_n$ or Π_n for some $n \Leftrightarrow A \leq_1 S^{(m)}$ for some $m \Leftrightarrow A \leq_T O^{(q)}$ for some q. This follows from the Lemma of § 1.

(4) Of the predicate hierarchy and the T-degrees, neither gives a refinement of the other. For Σ_n, $n > 0$, contains sets from more than one T-degree, and $O^{(n)}$, $n > 0$, contains sets in Σ_n but not in Π_n (e.g. $S^{(n)}$), in Π_n but not in Σ_n (e.g. $\overline{S^{(n)}}$), and in $\Sigma_{n+1} \cap \Pi_{n+1}$ but not in $\Sigma_n \cup \Pi_n$ (e.g. the set constructed in (5) of § 1). However, the 1-degrees (isomorphism types) give a common refinement of both.

Part II

§ 4. The sets

We shall locate certain sets in the above structures. Our goal in each case is to locate the 1-degree of a set. In most cases, this will be accomplished. We consider the following sets.

$$A_1 =_{\mathrm{df}} \{x \mid W_x \text{ is recursive}\}.$$
$$A_2 =_{\mathrm{df}} \{x \mid W_x \text{ is simple}\}.$$
$$A_3 =_{\mathrm{df}} \{x \mid W_x \text{ is hypersimple}\}.$$
$$A_4 =_{\mathrm{df}} \{x \mid W_x \text{ is creative}\}.$$
$$A_5 =_{\mathrm{df}} \{x \mid W_x \text{ is complete}\}.$$

The concepts used are defined in POST [3]. We restate them here.

W_x is *simple* if $\overline{W}_x$ is infinite but contains no infinite recursively enumerable subset.

W_x is *hypersimple* if $\overline{W}_x$ is infinite but there is no recursive function f such that for all n, $f(n) \geqq x_n$ where $\{x_n\}$ are the members of $\overline{W}_x$ in increasing order.

W_x is *creative* if $S^{(1)} \leqq_1 W_x$, i.e. if $W_x \equiv S^{(1)}$. (See MYHILL [12].)

W_x is *complete* if $S^{(1)} \leqq_T W_x$, i.e. if $W_x \in O^{(1)}$.

Results. We shall obtain that

$$A_1 \equiv A_4 \equiv S^{(3)},$$
$$A_2 \equiv A_3 \equiv \overline{S^{(3)}},$$
$$S^{(3)} \leqq_1 A_5 \leqq_1 S^{(4)}.$$

As a corollary to our work we shall also obtain that

$$\{x \mid \overline{W}_x \text{ is finite}\} \equiv \overline{S^{(3)}}.$$

Post's Problem is the question of whether or not $\overline{A}_1 = A_5$. Since $\overline{S^{(3)}} \equiv \overline{A}_1$, and $S^{(3)} \leqq_1 A_5$, but not $S^{(3)} \leqq_1 \overline{S^{(3)}}$, these results give as corollary the solution of Post's Problem: viz. $\overline{A}_1 \neq A_5$[1]).

Define: $A_6 =_{\mathrm{df}} \{x \mid$ for infinitely many y, $y \in W_x$ and W_y is infinite$\}$. A_6 will be discussed further at the conclusion of the paper. The exact classifications of A_5 and A_6 are interesting unsolved problems, as also is the question of relating A_5 to A_6. As we shall see, $\overline{S^{(3)}} \leqq_1 A_6 \leqq_1 \overline{S^{(4)}}$, and $S^{(3)} \leqq_1 A_6$.

Comment. The methods we describe can be used to locate sets with respect to $S^{(n)}$ and $\overline{S^{(n)}}$. As there are known to be, in addition to the $S^{(n)}$ and $\overline{S^{(n)}}$, many intermediate 1-degrees (see [4]), it cannot be expected that these methods will always yield an exact classification. It thus might be true that the result for A_5 is best possible by such methods.

We compute degrees by getting upper bounds and lower bounds. In § 5 we describe the technique for getting upper bounds; in § 6 we present the technique for lower bounds.

---

[1]) The above results for A_1, A_2, A_3 and A_4 were given in a paper presented to the Summer Institute in Mathematical Logic at Cornell in 1957. The result stated for A_4 was correct, but the suggested proof was erroneous. This is remedied in the present paper. Subsequently, in correspondence, Mostowski has informed the author that the results for A_1 and A_2 had been already obtained. Mostowski suggests a method for A_1 and A_2 that is simpler than that given below. Since the construction below for A_4 and A_5 depends upon that for A_1, I have not used the method of Mostowski.

§ 5. The Tarski-Kuratowski algorithm

Let α be a formula of lower predicate calculus with a single free individual variable. α can be transformed into an equivalent formula β in prenex form: $(Q_1 y_1) \ldots (Q_n y_n)$ [– – –], where [– – –] is a certain combination of the original predicate variables under the connectives of propositional calculus. If N is taken as domain, and if the predicate variables are interpreted as expressing certain given recursive relations, this provides, as is easily shown, an interpretation for [– – –] as a recursive relation. Hence we can interpret β as the predicate form $(Q_1 y_1) \ldots (Q_n y_n)\, R(y_1, \ldots, y_n, x)$ with recursive R, and, under the given interpretation, both α and β define the same arithmetical set. The procedure now to be described is simply: given any α, find β in such a way as to yield the lowest possible Σ_n, Π_n classification.

Given a set A, the procedure is therefore: (i) describe A in lower predicate calculus with recursive predicates; (ii) go to prenex form; (iii) collapse quantifiers of like kind; (iv) since, after (i), only a finite number of ways of doing (ii) and (iii) are possible, use those which give the shortest quantifier prefixes.

Applications of this procedure can be described by means of a special symbolism which we now outline.

Primitive symbols: $\forall\ \exists\ \Leftrightarrow\ \Rightarrow\ \&\ \vee\ \sim$ [ ].

Sample formulas: $[\Rightarrow]$, $\forall$, $\forall[\exists \Leftrightarrow \forall\exists]$, $[\exists \Rightarrow \exists]$.

Meaning: $[\Rightarrow]$ represents a formula of the form $[R_1(\ldots) \Rightarrow R_2(\ldots)]$ where R_1 and R_2 are recursive.

$\forall$ represents a formula of the form $(\forall x)\, R(\ldots x \ldots)$ where R is recursive.

. . . .

Transformation rules:

(i) rules for prenex form. E.g., starting from $\forall[\exists \Leftrightarrow \forall\exists]$, these rules could give in sequence: $\forall[\exists \Leftrightarrow \forall\exists]$; $\forall[[\exists \Rightarrow \forall\exists]\ \&\ [\forall\exists \Rightarrow \exists]]$; $\forall[\forall[\exists \Rightarrow \exists]\ \&$ $\&\ \exists[\forall\exists \Rightarrow]]$; $\forall\forall\exists[\exists\forall\ \&\ \exists\forall]$; $\forall\forall\exists\exists\exists\forall\forall$.

(ii) rules for collapsing like quantifiers. E.g. these would carry us from $\forall\forall\exists\exists\exists\forall\forall$ to $\forall\exists\forall$.

Thus, in accord with the above illustrative sequence, any set definable by a formula of the form $(\forall y_1)\, [(\exists y_2)\, R_1(y_1, y_2, x) \Leftrightarrow (\forall y_3)\, (\exists y_4)\, R_2(y_1, y_3, y_4, x)]$ where R_1 and R_2 are recursive is, in fact, in Π_3.

Similarly $[\exists \Rightarrow \exists]$ yields either $\forall\exists$ or $\exists\forall$, and such a set would be in $\Sigma_2 \cap \Pi_2$.

For a simple example, consider $A = \{x | W_x \text{ is infinite}\}$. Then: $x \in A$ if and only if $(\forall y)\, (\exists z)$ [z is a computation (under a suitable Gödel numbering of computations) from φ_x, and the output (value of φ_x) under z is greater than y]. The expression in brackets gives a recursive relation and we immediately have $A \in \Pi_2$.

Similarly take $B = \{x | \overline{W}_x \text{ is infinite}\}$. Then: $x \in A$ if and only if $(\forall x)\, (\exists y)$ $[y > x\ \&\ (\forall z)$ [z is not a computation from φ_x giving output y]]. Thus we have $\forall\exists[\&\forall]$. This gives $\forall\exists\forall$ and $B \in \Pi_3$.

For a more subtle example consider $A_1 = \{x | W_x$ is recursive$\}$. Our first task is to express A_1 in terms of quantifiers and recursive relations. Now $x \in A_1$ if and only if $(\exists y)\,[W_x = \overline{W}_y]$. This in turn gives: $x \in A_1$ if and only if $(\exists y)\,(\forall z)\,[z \in W_x \Leftrightarrow z \notin W_y]$. And this in turn is $(\exists y)\,(\forall z)\,[(\exists w)\,[w$ is a computation from φ_x with output $z] \Leftrightarrow \sim(\exists v)\,[v$ is a computation from φ_y with output $z]]$. Thus, going to our calculus, $\exists\forall[\exists \Leftrightarrow \sim\exists]$. This gives $\exists\forall[\exists \Leftrightarrow \forall]$; $\exists\forall[[\exists \Rightarrow \forall] \,\&\, [\forall \Rightarrow \exists]]$; $\exists\forall\forall\exists\exists$; $\exists\forall\exists$. Therefore $A_1 \in \Sigma_3$ and hence by (1) of § 3, we have

Lemma 1. $A_1 \leq_1 S^{(3)}$.

Similarly we obtain the following.

$A_2 = \{x | W_x$ is simple$\}$. By definition, $x \in A_2$ if and only if $\overline{W}_x$ infinite $\&(\forall y)\,[W_y$ infinite $\Rightarrow W_y \cap W_x \neq \theta]$. This becomes $\forall\exists\forall \,\&\, \forall[\forall\exists \Rightarrow \exists]$, or $\forall\exists\forall$. Therefore $A_2 \in \Pi_3$ and we have

Lemma 2. $A_2 \leq_1 \overline{S^{(3)}}$.

$A_3 = \{x | W_x$ is hypersimple$\}$. By definition, $x \in A_3$ if and only if $\overline{W}_x$ is infinite and no recursive function majorizes the successive members of $\overline{W}_x$. This becomes: $\overline{W}_x$ infinite and $(\forall z)\,[\varphi_z$ is a function $\Rightarrow (\exists y)\,(\exists v)\,(\exists w)\,[w$ is a computation for $\varphi_z(y) = v$ and $(\exists u)\,[u$ gives (under a suitable Gödel numbering) a list of at least $v - y + 2$ computations from φ_x giving distinct outputs $\leq v]]]$. To say that φ_z is a function is to say that $(\forall y)\,(\exists w)\,[w$ is a computation for $\varphi_z(y)]$. Thus $x \in A_3$ becomes $\forall\exists\forall \,\&\, \forall[\forall\exists \Rightarrow \exists\exists\exists[\&\exists]]$. This gives $\forall\exists\forall \,\&\, \forall\exists\forall$ and this, in turn, is $\forall\exists\forall$. Thus $A_3 \in \Pi_3$, and we have

Lemma 3. $A_3 \leq_1 \overline{S^{(3)}}$.

$A_4 = \{x | W_x$ is creative$\}$. By definition, $x \in A_4$ if and only if $S^{(1)} \leq_1 W_x$. Now $S^{(1)}$ is recursively enumerable. Let φ_{x_0} have $S^{(1)}$ as range. Then $S^{(1)} \leq_1 W_x$ if and only if $(\exists z)\,[\varphi_z$ is a one-one function $\&\, (\forall y)\,[y \in W_{x_0} \Leftrightarrow \varphi_z(y) \in W_x]]$. To say that φ_z is a one-one function is to say that $(\forall y_1)\,(\forall y_2)\,(\exists w_1)\,(\exists w_2)$ $[w_1$ is a computation for $\varphi_z(y_1)$ & w_2 is a computation for $\varphi_z(y_2)$ & & $[y_1 \neq y_2 \Rightarrow$ output of w_1 differs from output of $w_2]]$. Thus $x \in A_4$ becomes $\exists[\forall\forall\exists\exists \,\&\, \forall[\exists \Leftrightarrow \exists]]$. This yields $\exists\forall\exists$, and we have $A_4 \in \Sigma_3$. This gives

Lemma 4. $A_4 \leq_1 S^{(3)}$.

$A_5 = \{x | W_x$ is complete$\}$. By definition, $x \in A_5$ if and only if $S^{(1)} \leq_T W_x$. $x \in A_5$ becomes $(\exists x_1)\,(\exists x_2)\,[\varphi_{x_1}$ is a function and φ_{x_2} is a function and $(\forall z)$ $[[z \in S^{(1)} \Leftrightarrow \varphi_{x_1}(z)$ succeeds for $W_x] \,\&\, [z \notin S^{(1)} \Leftrightarrow \varphi_{x_2}(z)$ succeeds for $W_x]]]$. This yields $\exists\exists[\forall\exists \,\&\, \forall\exists \,\&\, \forall[[\exists \Leftrightarrow \exists[\exists \,\&\, \forall]] \,\&\, [\forall \Leftrightarrow \exists[\exists \,\&\, \forall]]]]$. This reduces to $\exists\forall\exists\forall$. Thus $A_5 \in \Sigma_4$, and we have

Lemma 5. $A_5 \leq_1 S^{(4)}$.

Similar analysis gives $A_6 \in \Pi_4$, and hence

Lemma 6. $A_6 \leq_1 \overline{S^{(4)}}$.

As we shall see, Lemmas 1—4 are best possible results. Whether Lemmas 5 and 6 are best possible is unknown. Our results below leave open the possibility that A_5 is in Σ_3. We shall note that A_6 is in neither Σ_3 nor Π_3; however, the possibility will remain open that A_6 is in both Σ_4 and Π_4, a stronger result than Lemma 6.

Remark. Various initial descriptions of a set, though theoretically equivalent, may yield results of varying strength. For an example, we return to the problem [in (5) of § 1] of showing a set in $\Sigma_{n+1} \cap \Pi_{n+1}$ but not in $\Sigma_n \cup \Pi_n$ $(n > 0)$. Taking $A \in \Sigma_n \dot{-} \Pi_n$, we defined $B =_{\mathrm{df}} \{x | (\exists y) [[x = 2y \,\&\, y \in A] \vee \vee [x = 2y + 1 \,\&\, y \in \bar{A}]]$. If B were in Σ_n, then both A and $\bar{A}$ would be recursively enumerable in $S^{(n-1)}$. Then $A \leq_T S^{(n-1)}$ and $A \in \Sigma_n \cap \Pi_n$ contrary to the choice of A. Similarly for B in Π_n. Therefore $B \notin \Sigma_n \cup \Pi_n$. To show $B \in \Sigma_{n+1} \cap \Pi_{n+1}$, we might carry out a computation similar to that in the examples immediately above. This, however, only yields $B \in \Sigma_{n+1} \cap \Pi_{n+2} = \Sigma_{n+1}$. To get the stronger result, we must use the different but equivalent description $B = \{x | (\exists y) [x = 2y \,\&\, y \in A] \vee [x \text{ is odd } \& \ (\forall y) [x = 2y + 1 \Rightarrow \Rightarrow y \in \bar{A}]]\}$.

The general procedure of this section was first used in similar classification problems in the theory of projective sets, and is related to work of Tarski and Kuratowski in that area (see [15]). We therefore speak of it as the *Tarski-Kuratowski algorithm.*

§ 6. Characterizing $S^{(n)}$

To find lower bounds, our method is to use certain rather easily and naturally defined sets isomorphic to $S^{(n)}$ (or to $\overline{S^{(n)}}$) and to show that one of these is $\leq_1$ the set being bounded.

We illustrate the method first with $S^{(2)}$. Consider $B =_{\mathrm{df}} \{x | W_x \text{ is infinite}\}$. This set is $\leq_1 \overline{S^{(2)}}$ by § 5. Now $\overline{S^{(2)}} \in \Pi_2$. Hence there is a recursive R such that $x \in \overline{S^{(2)}}$ if and only if $(\forall y)\,(\exists z)\, R(y, z, x)$. Given any x, consider the partial recursive function ψ defined by: $\psi(y) = y$ if $[(\exists z) R(y,z,x) \,\&\, (\forall v)\ [v < y \Rightarrow \psi(v)$ is defined$]]$. An index for ψ is uniformly effectively computable from x, call it $h(x)$ with h a recursive function. Clearly $x \in \overline{S^{(2)}} \Leftrightarrow W_{h(x)}$ is infinite. As we shall see in a moment, h can be easily rendered a one-one recursive function. This gives $\overline{S^{(2)}} \leq_1 B$, and hence $B \equiv \overline{S^{(2)}}$.

Using B as a standard reference, we can now classify other sets. Thus take $C =_{\mathrm{df}} \{x | \varphi_x$ is a recursive permutation$\}$. The algorithm of § 5 shows $C \in \Pi_2$. Hence $C \leq_1 \overline{S^{(2)}}$. If we could show $B \leq_1 C$, we would have $C \equiv \overline{S^{(2)}}$ and thus know the 1-degree of C. Let x be given. Define ψ by: $\psi(z) = z$ if $(\mathsf{E}y)$ $[y \in W_x \,\&\, y > z]$. An index for ψ is clearly computable from x in a uniform effective manner; call it $g(x)$ with g a recursive function. Then $x \in B \Leftrightarrow g(x) \in C$. It remains to show that g can be made one-one. This is done by use of a recursive function of two variables t with the properties: (i) $t(x, y) = t(x', y') \Rightarrow \Rightarrow [x = x' \,\&\, y = y']$; and (ii) $\varphi_{t(x,y)} = \varphi_x$ for all x and y. Given such a t, $g'(x) = t(g(x), x)$ gives a one-one recursive function such that $x \in B \Leftrightarrow g'(x) \in C$. Thus $B \leq_1 C$ and $C \equiv B \equiv \overline{S^{(2)}}$. In the same way the function h of the preceding paragraph can be rendered one-one. The function t, suggested by Davis, is defined formally by getting first a function $t'(x, y)$ with the properties: (i) $t'(x, y) = t'(x', y') \Rightarrow [y = y' \vee x \neq x']$; and (ii) $\varphi_{t'(x,y)} = \varphi_x$ for all x and y. (t' is obtained, e.g., by taking $t'(x, y)$, as y varies, to be Gödel numbers of longer and longer redundant definitions for the fixed partial recursive function

$\varphi_{x'}$) t is then defined by $t(x, y) = t'(x, z)$ where $z = \mu w[t'(x, w) \neq t(x', y')$ for all x', y' such that $2^{x'}3^{y'} < 2^x 3^y]$.

By the same method as used above for C, with the standard sets B and $\overline{B}$ used to get lower bounds, many of the results of [7] can be obtained.

Our results below will concern the three quantifier case. As a standard set, we take

$$D =_{\text{df}} \{x | (\exists y) [y \in W_x \,\&\, W_y \text{ is infinite}]\} .$$

We first show that $D \equiv S^{(3)}$.

Theorem 1. $D \equiv S^{(3)}$.

Proof. By the Tarski-Kuratowski algorithm, $D \in \Sigma_3$, and therefore $D \leqq_1 S^{(3)}$.

Conversely, let $S^{(3)}$ be defined by $(\exists y_1)$ $(\forall y_2)$ $(\exists y_3)$ $R(y_1, y_2, y_3, x)$. Define a recursive function ψ as follows. $\psi(y_1)$ is an index for the partial recursive function θ such that, as y_2 varies, $[\theta(y_2) = y_2$ if $[(\exists y_3)$ $R(y_1, y_2, y_3, x)$ & $(\forall v)$ $[v < y_2 \Rightarrow \theta(v)$ defined$]]$ and $\theta(y_2)$ undefined otherwise$]$. Let h be a recursive function giving $h(x)$ as an index for ψ. Then $x \in S^{(3)} \Leftrightarrow h(x) \in D$. By means of the t function, h can be made a one-one function. Hence $S^{(3)} \leqq_1 D$, and this completes the proof.

We next proceed to the set A_1. We shall find a recursive function h such that given any x, $x \in D \Rightarrow W_{h(x)}$ has finite complement (and is therefore a recursive set), while $x \in \overline{D} \Rightarrow W_{h(x)}$ has non-recursively enumerable complement (and is therefore not a recursive set). Thus $x \in D \Leftrightarrow h(x) \in A_1$. Setting $h'(x) = t(h(x), x)$ gives h' a one-one recursive function such that $x \in D \Leftrightarrow h'(x) \in A_1$, and we can conclude $D \leqq_1 A_1$. By Lemma 1 this yields

Theorem 2. $A_1 \equiv S^{(3)}$.

From an illustrative example used in § 5, this construction also gives a corollary.

Corollary. $\{x | \overline{W}_x \text{ finite}\} \equiv S^{(3)}$.

For a proof of this theorem and corollary it remains to describe h. It will be enough to show how, given any x, we can go effectively to instructions for enumerating $W_{h(x)}$. We give the construction in somewhat anthropomorphic terms, believing that the effectiveness of the method is most clearly and directly presented in this way. (A more formal treatment of a very similar construction is given in [6].)

We take the integers as given to us in a single verticle list, which we call the *main list*. As the procedure goes on, we shall write a *plus* beside certain integers in this list. The integers receiving this mark will constitute $W_{h(x)}$. We first describe several procedures which will serve as successive approximations to the ultimate correct procedure.

The *preliminary procedure* is as follows. Let k be any constant > 0. Let $a_0, a_1, \ldots$ be an effectively enumerable infinite sequence of distinct integers.

Stage 1. The set W_0 is associated with the integer a_0 in the main list. A computation of W_0 is begun and carried through k steps. If this produces the integer a_0 as a member of W_0, we write a *plus* beside a_0 in the main list.

Stage 2. The set W_1 is associated with the integer a_1 in the main list. Computations for W_0 and W_1 are begun and each carried through $2k$ steps.

If this produces a_0 as a member of W_0, we write a *plus* beside a_0 in the main list if a *plus* is not already there. If this produces a_1 as a member of W_1, we write a *plus* beside a_1 in the main list.

Stage 3. . . . . . [similarly, with W_2 associated with a_2 and computation for each of W_0, W_1 and W_2 carried through $3k$ steps] . . . .

. . . .

Note that this preliminary procedure gives, by its *plus* marks, a non-recursive set; for the complement of this set differs from every recursively enumerable set, since for every n, $a_n \in W_n \Leftrightarrow a_n$ is not in that complement. (If $a_0, a_1, \ldots$ is the sequence $0, 1, 2, \ldots$, the preliminary procedure gives the set K of [3].) In the preliminary procedure, we call a_n the *position* of W_n. As the procedure goes on, any integer of the main list which has not yet been used as a position for a W_n will be called *free*, and any integer of the main list which does not yet have a *plus* beside it will be called *vacant*. The notions of *position*, *free* and *vacant* will also be used, in an obvious way, in connection with certain modified forms of the preliminary procedure.

Next we describe an *auxiliary procedure* which is performed as follows. It is separate from and does not concern the main list. Given an x, we begin to list in a horizontal row the distinct members of W_x. Call these $y_0, y_1, y_2, \ldots$. Under each number y_i, we also begin to list the distinct members of W_{y_i} in a vertical column. This auxiliary procedure is performed in stages as follows.

Stage 1. Perform k steps in the enumeration of W_x. Take each member $y \in W_x$ obtained and perform k steps in the enumeration of W_y.

Stage 2. Perform $2k$ steps in the enumeration of W_x. Take each member $y \in W_x$ obtained and perform $2k$ steps in the enumeration of W_y.

. . . .

We now go to our *final procedure* by using the auxiliary procedure to make certain modifications in the preliminary procedure. Roughly speaking, we shall begin the preliminary procedure as previously described, but shall then change the positions of certain W_n according as certain events occur in the auxiliary procedure. In the case of each change, a W_n being given a new position will be associated with a previously free integer. We now describe the final procedure. Let any x be given.

Stage 1. Perform stage 1 of the preliminary procedure with $a_0 = 0$.

Stage 2. Perform stage 1 of the auxiliary procedure using the given x. If 0 appears in W_x and if any values of W_0 appear in the column of that 0, change the position of W_0 to 1 on the main list, and write a *plus* beside 0 on the main list.

Stage 3. Associate W_1 with the least free integer (i.e. not previously associated with W_0). Let $a_0^{(1)}$ and $a_1^{(1)}$ be the positions of W_0 and W_1 respectively. Compute $2k$ steps in the enumeration of each of W_0 and W_1. If $a_0^{(1)}$ appears in W_0 and $a_0^{(1)}$ is vacant, put a *plus* beside $a_0^{(1)}$ in the main list. Similarly for $a_1^{(1)}$ and W_1.

Stage 4. Perform stage 2 of the auxiliary procedure. If 0 appears in W_x and if any new values appear under 0 (i.e. values that have not appeared in earlier stages of the auxiliary procedure), then put a *plus* beside each of $a_0^{(1)}$ and $a_1^{(1)}$ on the main list and move W_0 and W_1 down to the least two free integers on the main list. If this does not occur, but if a 1 appears in W_x and any new values appear under 1, then put a *plus* beside $a_1^{(1)}$ and move W_1 down to the least free integer.

. . . .

Stage $2n + 1$. Associate W_n with the least free integer. Let $a_0^{(n)}, a_1^{(n)}, \ldots, a_n^{(n)}$ be the current positions of $W_0, \ldots, W_n$ respectively. Compute $(n + 1)k$ steps in the enumeration of each of $W_0, W_1, \ldots, W_n$. If $a_i^{(n)}$ appears in W_i for any $0 \leqq i \leqq n$, and if $a_i^{(n)}$ is vacant, put a *plus* beside $a_i^{(n)}$ in the main list.

Stage $2n + 2$. Perform stage $n + 1$ of the auxiliary procedure. If any new members appear under any $y \in W_x \cap \{0, 1, 2, \ldots, n\}$, take the least such y, call it m. (By new members under y, we mean numbers not listed under y in previous stages of the auxiliary procedure.) Put a *plus* beside $a_i^{(n)}$ for all vacant $a_i^{(n)}$ such that $m \leqq i \leqq n$, and then change the positions of W_i, $m \leqq i \leqq n$, to the least $n - m + 1$ free integers in the main list.

. . . .

The integers in the main list receiving a *plus* constitute $W_{h(x)}$. If $x \in \overline{D}$, then W_i for each i is moved only finitely often. Let z_i be the final position of W_i, for all i. Then $z_i \in W_i \Leftrightarrow z_i \in W_{h(x)}$. Hence $\overline{W}_{h(x)}$ is not recursively enumerable and hence $W_{h(x)}$ is not recursive; i.e. $h(x) \in \overline{A}_1$. On the other hand, if $x \in D$, then W_i, for some i, will be moved infinitely often, and all integers below some point in the main list will have a *plus*. Hence $W_{h(x)}$ will have finite complement, and $h(x) \in A_1$. This completes the proof.

In [16], DEKKER gives a uniform effective method for going from a recursively enumerable set W_x to a corresponding recursively enumerable set H_x such that if W_x is not recursive then H_x is hypersimple. The method is as follows. Let $z_0, z_1, \ldots$ be distinct members in the effective enumeration of W_x given by index x. Define $H_x = \{n | (\exists m) [m > n \;\&\; z_m$ appears in the enumeration of $W_x \& z_m < z_n]\}$. As can be verified directly from the definition of hypersimple set in § 4, H_x is hypersimple if W_x is not recursive. We here note further that H_x is recursive if W_x is recursive and that the method gives a recursive function g such that $H_x = W_{g(x)}$. Hence we have $x \in \overline{A}_1 \Leftrightarrow g(x) \in A_3$ and $x \in \overline{A}_1 \Leftrightarrow g(x) \in A_2$, (since every hypersimple set is simple and since a recursive set can be neither simple nor hypersimple). Now, using the function t, we have $x \in \overline{A}_1 \Leftrightarrow t(g(x), x) \in A_3 \Leftrightarrow t(g(x), x) \in A_2$; and this gives $\overline{A}_1 \leqq_1 A_2$ and $\overline{A}_1 \leqq_1 A_3$. Therefore, using Lemmas 2 and 3, we have

Theorem 3. $A_2 \equiv A_3 \equiv \overline{S}^{(3)}$.

To locate the set A_4, we prove that $D \leqq_1 A_4$. We do this by giving a recursive function h such that if $x \in D$ then $W_{h(x)}$ is creative, and if $x \in \overline{D}$ then $W_{h(x)}$ is not complete. Since A creative $\Rightarrow A$ complete, this also gives the corollary

that $D \leq_1 A_5$. These results, together with Lemmas 4 and 5 from § 5, give us our final theorems.

Theorem 4. $A_4 \equiv S^{(3)}$.

Theorem 5. $S^{(3)} \leq_1 A_5 \leq_1 S^{(4)}$.

To describe h, we give a uniform method showing how to list $W_{h(x)}$ when x is given. As in the case of A_1, we shall reach the method by successive approximations.

First we describe a preliminary procedure. (As the reader will observe, it is a somewhat more complex variety of the previous preliminary procedure.) We begin with two main lists, each consisting of 0, 1, 2, 3, ... Our procedure will lead us to write a *plus* beside certain integers in each list. We shall thus obtain two recursively enumerable sets, one from each list. Call these sets B and C. We call the first main list the *B-list*, the second the *C-list*. We describe the procedure in stages. The finite set of integers in the B-list receiving *plus* marks by stage n we call B_n; and the finite set of integers in the C-list receiving *plus* marks by stage n we call C_n. $B_0 = C_0 = \emptyset$. Clearly $B_0 \subset B_1 \subset B_2 \subset \cdots$, and $C_0 \subset C_1 \subset C_2 \subset \cdots$, and $B = \bigcup_{n=0}^{\infty} B_n$ and $C = \bigcup_{n=0}^{\infty} C_n$. Let k be some constant > 0.

Stage 1. Associate W_0^X (X as yet unspecified) with 0 in the B-list. Compute k steps in the computation of $W_0^{C_0}$. If 0 appears, put a *plus* beside 0 in the B-list. In general, the computation of 0 from $W_0^{C_0}$, if it occurs, will depend upon the absence of certain (finitely many) integers from C_0. Put a *minus* beside these integers in the C-list.

Stage 2. Associate the set W_0^Y (Y as yet unspecified) with the first integer c_0 in the C-list which has no minus beside it or below it. Compute k steps in the enumeration of $W_0^{B_1}$. If c_0 appears, put a *plus* beside it in the C-list and a *minus* beside those integers in the B-list whose absence from B_1 was used.

We call an integer in one of the lists *free* (at a given stage) if neither it nor any integer below it in that list has a *plus*, a *minus*, a W_i^X or a W_i^Y (for any i) associated. We call an integer in a list *vacant* if it does not have a *plus* beside it.

. . . .

Stage $2n+1$. Associate W_n^X with the first free integer in the B-list. Let $b_0^{(n)}, \ldots, b_n^{(n)}$ be the positions of $W_0^X, \ldots, W_n^X$. Compute $(n+1)k$ steps in each of $W_0^{C_{2n}}, \ldots, W_n^{C_{2n}}$. Let $b_j^{(n)}$ be the least member of $\{b_0^{(n)}, \ldots, b_n^{(n)}\}$ such that $b_j^{(n)}$ is vacant and $b_j^{(n)}$ appears in the computation of $W_j^{C_{2n}}$ just made. (If there is no such $b_j^{(n)}$, go on to stage $2n+2$.) Put a *plus* beside $b_j^{(n)}$ in the B-list and a *minus* beside the integers in the C-list whose vacancy was used in getting $b_j^{(n)}$ in $W_j^{C_{2n}}$. Then, if $j < n$, go to the C-list and change positions by moving the W_i^Y, $j \leq i < n$, down to free integers in the C-list.

Stage $2n+2$. Associate W_n^Y with the first free integer in the C-list. Let $c_0^{(n)}, \ldots, c_n^{(n)}$ be the positions of $W_0^Y, \ldots, W_n^Y$. Compute $(n+1)k$ steps in each of $W_0^{B_{2n+1}}, \ldots, W_n^{B_{2n+1}}$. Let $c_j^{(n)}$ be the least member of $\{c_0^{(n)}, \ldots, c_n^{(n)}\}$ such that $c_j^{(n)}$ is vacant and $c_j^{(n)}$ appears in the computation of $W_j^{B_{2n+1}}$ just made. (If there is no such $c_j^{(n)}$, go on to stage $2n+3$.) Put a *plus* beside $c_j^{(n)}$ in the

C-list and a *minus* beside the integers in the B-list whose vacancy was used in getting $c_j^{(n)}$ in $W_j^{B_{2n+1}}$. Then, if $j < n$, go to the B-list and change positions by moving the W_i^X, $j < i \leqq n$, down to free integers in the B-list.

. . . .

(Note that it is possible for an integer in either list to have first a *minus* and then a *plus* beside it.)

The reader will observe (by a simple inductive proof) that each W_i^X and each W_i^Y gets moved only finitely often. (W_0^X is never moved. W_0^Y can be moved at most once. W_1^X can be moved at most once for each position of W_0^Y, etc.) The reader will also observe that if b_i is the final position of W_i^X, then b_i gets a *plus* beside it if and only if $b_i \in W_i^C$, (since the computation of b_i in W_i^C depends only on the existence of certain finite sets of integers in C and in $\bar{C}$). Hence, for all i, $b_i \in B \Leftrightarrow b_i \in W_i^C$; hence $\bar{B}$ is not recursively enumerable in C; hence $B \nleqq_T C$. Similarly $C \nleqq_T B$. Thus both B and C are non-recursive incomplete sets. The preliminary procedure just described is, with slight modification for our purposes, the construction of FRIEDBERG that solves Post's Problem.

Next, as auxiliary procedure, we take the same auxiliary procedure as was used before in the proof of Theorem 2.

We now use this auxiliary procedure to modify the preliminary procedure in the following way. As C-list we use the integers in order as before. For our B-list, we make the following tabular arrangement of the integers. In a first column we list, in increasing order, all integers which are not positive powers of primes (viz. 0, 1, 6, 10, . . .). In the second column we list the successive positive powers of 2, in the third column the powers of 3, and in general, in the $n + 2^{nd}$ column, the successive powers of p_n, where p_n is the $n + 1^{st}$ prime. For the C-list, our procedure is much as in the preliminary procedure. In the case of the B-list, we associate the W_i^X only with integers in the first column. A further modification is that when *minus* is entered by any integer in any column of the B-list, because of the appearance of a *plus* in the C-list for some W_j^Y, we index the minus with j; that is, we write "$-_j$". We call this last symbol a *j-minus*. Finally, we use the auxiliary procedure, much as in the case of Theorem 2, to cause additional downward shifts of both the W_i^X and the W_i^Y, and at the same time we use it to bring about the appearance of *plus* beside certain integers in columns 2, 3, . . . of the B-list. This is described below. The final procedure is described in stages. Sets $B_0, B_1, \ldots, C_0, C_1, \ldots$ are defined just as in the case of the preliminary procedure. Let K be a fixed creative set (e.g. $\{z|z \in W_z\}$), and let $z_0, z_1, z_2, \ldots$ be the distinct elements of K in some effective enumeration. As before, k is some constant > 0. Let any x be given.

. . . .

Stage $3n + 1$. Associate W_n^X with the first free integer in column one of the B-list. Let $b_0^{(n)}, \ldots, b_n^{(n)}$ be the . . . [Remainder of description exactly as in stage $2n + 1$ of the preliminary procedure except that '$2n$' is replaced by '$3n$' wherever it occurs].

Stage $3n+2$. [Description exactly as in stage $2n+2$ of the preliminary procedure except that '$2n$' is replaced by '$3n$' wherever it occurs, that '*minus*' is replaced by '*j-minus*', and that the last occurrence of '*B*-list' is replaced by 'column one of the *B*-list'.]

Stage $3n+3$. Perform stage $n+1$ of the auxiliary procedure. If any new numbers appear under any $y \in W_x \cap \{0, 1, \ldots, n\}$, take the least such y, call it m. (By new numbers under y, we mean numbers not listed under y in previous stages of the auxiliary procedure.) Then, if such an m exists, do the following (otherwise go on to stage $3n+4$): (i) move all W_i^X, $m \leq i \leq n$, down to free integers in column one of the *B*-list; (ii) move all W_i, $m \leq i \leq n$, down to free integers in the *C*-list; (iii) go to the column of powers of p_m, and in this column put *plus* beside all vacant members of the set $\{p_m^{z_0+1}, p_m^{z_1+1}, \ldots, p_m^{z_n+1}\}$ with the proviso that any *j-minus* occurring by a member of this set is replaced by a *plus* if $m \leq j$, but that any *j-minus* is left alone and a *plus* is not used if $j < m$.

. . . .

The set B is now taken to be the desired $W_{h(x)}$. Note first that if $x \in \bar{D}$, then by a similar inductive argument to that indicated after the preliminary procedure, each W_i^X and W_i^Y is moved only a finite number of times. Hence, as before, $B \not\leq_T C$ and $C \not\leq_T B$. Therefore $B = W_{h(x)}$ is not complete. On the other hand, if $x \in D$, then for some y in W_x, y has an infinite column in the auxiliary procedure. Let q be the least such y. Then W_i^X and W_i^Y, $i < q$, move only finitely often, but W_i^X and W_i^Y, $q \leq i$, move infinitely often. Hence only finitely many occurrences of *j-minus*, $j < q$, appear in the *B*-list. Hence $p_q^{z+1} \in B$ for all but a finite number of integers z in K. Furthermore, by the construction, if $z \in \bar{K}$, then $p_q^{z+1} \in \bar{B}$. Let $\hat{B} = \{p_q^{z+1} | z \in K\}$. Then $B \cup \hat{B}$ differs from B by aa finite set. Since $z \in K \Leftrightarrow p_q^{z+1} \in B \cup \hat{B}$, we have $K \leq_1 B \cup \hat{B}$. Therefore, by the definition in § 4, $B \cup \hat{B}$ is creative. Therefore B is creative, since, as is easily shown, a creative set remains creative when a finite set is deleted from it. Thus if $x \in D$, $W_{h(x)}$ is creative. Theorems 4 and 5 now follow.

Note that this result gives the solution to Post's Problem as a corollary. Since Friedberg's construction is used in the proof, this is hardly surprising. It does give, however, the stronger result that A_1 and $\bar{A}_5$ are of different 1-degree of unsolvability; an investigation of these degrees was suggested by Davis and Shapiro as a possible approach to Post's Problem.

That A_1 and $\bar{A}_5$ are of different degree follows from the fact that $A_1 \leq_1 A_5$ (since $A_1 \equiv S^{(3)}$ and $S^{(3)} \leq_1 A_5$). For $A_1 \equiv \bar{A}_5$ would imply $A_1 \leq_1 \bar{A}_1$, which would imply $S^{(3)} \leq_1 \overline{S^{(3)}}$, which would imply (taking complements) $S^{(3)} \equiv \overline{S^{(3)}}$; since $S^{(3)}$ is recursively enumerable in $S^{(2)}$, this would give $S^{(3)} \leq_T S^{(2)}$, a contradiction.

The question of where to place A_5 between $S^{(3)}$ and $S^{(4)}$ remains open.

Remark. As is well known, or else is shown above,

$S^{(1)} \equiv \{x | (\exists y) [y \in W_x]\}$;

$S^{(2)} \equiv \{x | W_x \text{ is finite}\}$;

$S^{(3)} \equiv \{x | (\exists y)\ [y \in W_x \text{ and } W_y \text{ is infinite}]\}$. This suggests the possibility, $S^{(4)} \equiv \bar{A}_6 = \{x |$ for only finitely many $y \in W_x$, W_y is infinite$\}$.
If this suggested isomorphism and generalizations of it were true, it would give an interesting and useful sequence of characterizations of $S^{(n)}$[2]). (The set, $\{x |$ for infinitely many $y \in W_x$, W_y is finite$\}$ can be shown $\equiv S^{(3)}$ without difficulty.)

It is not difficult to show that $S^{(3)} \leqq_1 \bar{A}_6 \leqq_1 S^{(4)}$ and $\overline{S^{(3)}} \leqq_1 \bar{A}_6$, but the writer has been unable to get a stronger result, despite some effort. $\bar{A}_6 \leqq_1 S^{(4)}$ is true by Lemma 6. $\overline{S^{(3)}} \leqq_1 \bar{A}_6$ is straightforward. To show $S^{(3)} \leqq_1 \bar{A}_6$, the construction is roughly as follows. We want an f such that $x \in D \Leftrightarrow f(x) \in \bar{A}_6$. Given x, we construct $W_{f(x)}$ to be an infinite set with each member of $W_{f(x)}$ the index for a set which is to be infinite except for possible truncation by operation of the auxiliary procedure of Theorem 2. Each time a new member appears in the auxiliary procedure, this truncates a certain finite collection of sets with indices in $W_{f(x)}$. It is not difficult to arrange this so that any infinite column in the auxiliary procedure acts to truncate all but a finite number of the indices in $W_{f(x)}$, and so that if all columns of the auxiliary procedure are finite, then infinitely many indices of infinite sets appear in $W_{f(x)}$. This then yields $S^{(3)} \leqq_1 \bar{A}_6$.

The writer conjectures that $A_5 \equiv \bar{A}_6 \equiv S^{(4)}$. A first step toward verifying this conjecture would be to show that $\overline{S^{(3)}} \leqq_1 A_5$.

References

[1] Kleene, S.: Recursive predicates and quantifiers. Trans. Amer. math. Soc. **53**, 41—73 (1943). — [2] Mostowski, A.: On definable sets of positive integers. Fund. math. **34**, 81—112 (1947). — [3] Post, E.: Recursively enumerable sets of positive integers and their decision problems. Bull. Amer. Math. Soc. **50**, 284—316 (1944). — [4] Kleene, S., and E. Post: The upper semi-lattice of degrees of recursive unsolvability. Ann. of Math. ser. 2, **59**, 379—407 (1954). — [5] Turing, A.: On computable numbers with an application to the Entscheidungsproblem. Proc. London math. Soc., ser. 2, **42**, 230—265 (1936/37). — [6] Friedberg, R.: Two recursively enumerable sets of incomparable degrees of unsolvability. Proc. Nat. Acad. Sci. (Wash.) **43**, 236—238 (1957). — [7] Mostowski, A.: Examples of sets definable by means of two and three quantifiers. Fund. math. **42**, 259 bis 270 (1955). — [8] Robinson, R.: Arithmetical representation of recursively enumerable sets. J. Symb. Log. **21**, 162—186 (1956). — [9] Davis, M.: Computability and Unsolvability. New York: McGraw-Hill 1958. — [10] Rogers, H.: Gödel numberings of partial recursive functions. J. Symb. Log. **23**, 331—341 (1958). — [11] Friedberg, R., and H. Rogers: Reducibility and completeness for sets of integers. Z. math. Log. **5**, H. 2 (1959). — [12] Myhill, J.: Creative sets. Z. math. Log. **1**, 97—108 (1955). — [13] Spector, C.: On degrees of resursive unsolvability. Ann. of Math., ser. 2, **64**, 581—592 (1956). — [14] Shapiro, N.: Degrees of computability. Trans. Amer. math. Soc. **82**, 281—299 (1956). — [15] Tarski, A.: Sur les ensembles définissables de nombres réels I. Fund. math. **17**, 210—239 (1931). — [16] Dekker, J.: A theorem on hypersimple sets. Proc. Amer. math. Soc. **5**, 791—796 (1954).

(Eingegangen am 19. Januar 1959)

---

[2]) *Added in proof*, 23 *May* 1959. The generalizations suggested have been demonstrated by Kreisel, Shoenfield and Wang in work that will appear in a forthcoming paper by them. It follows that $\bar{A}_6 \equiv S^{(4)}$, and this answers one of the open questions raised below.

SCHÄFFER, J. J.
Math. Annalen 138, 141—144 (1959)

Addendum: Function Spaces with Translations

By

JUAN JORGE SCHÄFFER in Montevideo (Uruguay)

This note supplements the author's paper [1], which will be assumed to be known and shall be referred to in the sequel as F, its contents being quoted as, e. g., Theorem F.4.2, formula F(4.1). The present results, besides their intrinsic interest, have some significant applications to the theory of linear differential equations, which will be discussed in a forthcoming paper.

Let $\mathbf{F}$ be any space in the class $\mathcal{F}$. By Corollary F.4.3 there exists, for every $l > 0$, a number $\alpha(l) = \alpha(\mathbf{F}; l) > 0$ such that

$$\int_{J'} |\varphi(t)|\, dt \leq \alpha(l)\, |\varphi|_{\mathbf{F}} \tag{1}$$

holds for every interval $J' \subset J$ with $\mu(J') = l$ and every $\varphi \in \mathbf{F}$, and such that $\alpha(l)$ cannot be replaced in (1) by any smaller number if $J' \subset [\theta, \infty)$, where $\theta = \theta(\mathbf{F})$.

It is an interesting problem to find, for given $l > 0$ and every $J' \subset [\theta, \infty)$ with $\mu(J') = l$, a non-trivial function $\varphi \in \mathbf{F}$ which satisfies (1) with equality, or at least such that

$$\int_{J'} |\varphi(t)|\, dt \geq \lambda\, \alpha(l)\, |\varphi|_{\mathbf{F}}$$

for some fixed λ, $0 < \lambda < 1$. One of our results will be that, if $\mathbf{F}$ is locally closed and $\lambda = 1/2$, the characteristic function $\chi_{J'}$ (which we already know to belong to $\mathbf{F}$ by Theorems F. 4.7 and F. 4.19) has this property, that is,

$$|\chi_{J'}|_{\mathbf{F}} \leq 2l/\alpha(l) \tag{2}$$

(Theorem 2). Using (1), (2) may be rewritten as

$$|\chi_{J'}| \int_{J'} |\varphi(t)|\, dt \leq 2l\, |\varphi|_{\mathbf{F}} \tag{3}$$

for all $\varphi \in \mathbf{F}$. The proof of Theorem 2 contains a new proof of the fact that $\chi_{J'} \in \mathbf{F}$, independent of Theorem F. 4.19, which rests on the theory of associate spaces.

A space $\mathbf{F} \in \mathcal{F}$ which is not (quasi) locally closed need not contain the characteristic function of any interval; if it does, however, it contains the characteristic function of every bounded interval $J' \subset [\theta + \theta', \infty)$, where $\theta' \geq 0$. Since the characteristic functions of two intervals of the same length have the same norm in $\mathbf{F}$ as soon as both belong to $\mathbf{F}$ (by property (d)), we might conjecture that, in analogy with (3), we have

$$|\chi_{J''}|_{\mathbf{F}} \int_{J'} |\varphi(t)|\, dt \leq 2l\, |\varphi|_{\mathbf{F}} \tag{4}$$

for all intervals J', $J'' \subset J$ such that $\mu(J') = \mu(J'') = l$ and $\chi_{J''} \in \mathbf{F}$. This conjecture is false, even if $\mathbf{F}$ is complete and belongs to $\mathcal{F}^{\#}$ and even if the

constant 2 is replaced by any larger constant in (4), as Example 1 shows (but see Corollary 1 below). However, (4) does hold for all $\varphi \in \mathbf{F}$ which are *continuous in the closed interval* J' (Theorem 1).

Example 1. Let $\mathbf{D}$ be the space defined in Example F. 4.4 (the subspace of $\mathbf{L}^\infty$ spanned by $\mathbf{D}_0$, the linear manifold of all functions in $\mathbf{L}^\infty$ which vanish essentially except on a non-dense set). Let $\mathbf{T}$ be the space defined in F, Section 4.4. Both $\mathbf{D}$ and $\mathbf{T}$ are complete and belong to $\mathscr{T}^\#$; hence $\mathbf{F} = \mathbf{D} \vee \mathbf{T}$ shares these properties (Theorems F. 2.1 and F. 4.2) but is not quasi locally closed, as is easily seen.

Set $J' = [0, n]$; we shall prove that

$$|\chi_{J'}|_{\mathbf{F}} = n\ ; \tag{5}$$

indeed, $\chi_{J'} = \sum_0^{n-1} T_i^+ \chi_{[0,1]}$, whence $\chi_{J'} \in \mathbf{T}$ and $|\chi_{J'}|_{\mathbf{F}} \leqq |\chi_{J'}|_{\mathbf{T}} \leqq n$. It remains to prove the reverse inequality. For any bounded interval $J'' \subset J$ and any $\varphi \in \mathbf{D}_0$, φ vanishes a. e. on some subinterval of J'', whence $|\chi_{J''} - \varphi|_\infty \geqq 1$; since $\mathbf{D}_0$ is $\mathbf{L}^\infty$-dense in $\mathbf{D}$, this inequality persists for any $\varphi \in \mathbf{D}$. Assume that $\chi_{J'} = \varphi + \psi$, $\varphi \in \mathbf{D}$, $\psi \in \mathbf{T}$; it follows that $\operatorname{ess\,sup}_{t \in J''} |\psi(t)| \geqq 1$ for every interval $J'' \subset J'$. Now the space $\mathbf{S}$ (F, Section 4.4) is isometrically embedded in $\mathbf{T}$ (Theorem F. 4.13) and obviously every function in $\mathbf{T}$ which vanishes outside a bounded set also belongs to $\mathbf{S}$; in particular, $\psi' = \chi_{J'}\psi \in \mathbf{S}$, and $\operatorname{ess\,sup}_{t \in J''} |\psi'(t)| \geqq 1$ for every bounded interval $J'' \subset J'$. It follows from Lemma F. 4.5 that $|\psi|_{\mathbf{T}} \geqq |\psi'|_{\mathbf{T}} = |\psi'|_{\mathbf{S}} \geqq n$. Thus $|\varphi|_{\mathbf{D}} + |\psi|_{\mathbf{T}} \geqq n$ for all choices of φ, ψ, whence $|\chi_J|_{\mathbf{F}} \geqq n$, and this completes the proof of (5).

Let $E \subset J'$ be a non-dense set such that, say, $\mu(E) \geqq n/2$. Then $\chi_E \in \mathbf{D}$ and $|\chi_E|_{\mathbf{F}} \leqq |\chi_E|_{\mathbf{D}} = |\chi_E|_\infty = 1$. Hence

$$\frac{n\,|\chi_E|_{\mathbf{F}}}{|\chi_{J'}|_{\mathbf{F}} \int_{J'} \chi_E(t)\,dt} \leqq \frac{2}{n},$$

which can be made arbitrarily small if n is sufficiently large. Thus (4) fails, with $\varphi = \chi_E$, even if 2 is replaced by any larger (but fixed) constant.

Theorem 1. *If* $\mathbf{F} \in \mathscr{T}$ *contains the characteristic function of an interval, and if* $\varphi \in \mathbf{F}$ *is continuous on the closed bounded interval* $J' \subset J$, $\mu(J') = l > 0$, *then* (4) *holds, where* $J'' \subset J$ *is any interval such that* $\mu(J'') = l$ *and* $\chi_{J''} \in \mathbf{F}$.

Proof. Set $J' = [t_0, t_0 + l]$. We may assume that $\varphi(t)$ does not vanish identically on J'. Set $M = \max_{t \in J'} |\varphi(t)| > 0$, $C = \int_{J'} |\varphi(t)|\,dt > 0$. Let ε, $0 < \varepsilon < 1$ be given and set $\delta = C\varepsilon/2M$. Since $C \leqq Ml$, we have $2\delta \leqq l\varepsilon \leqq l$. We define $\lambda(t)$ by

$$\lambda(t) = \begin{cases} 0 & 0 \leqq t \leqq t_0 \\ (t - t_0)/\delta & t_0 \leqq t \leqq t_0 + \delta \\ 1 & t_0 + \delta \leqq t \leqq t_0 + l - \delta \\ (t_0 + l - t)/\delta & t_0 + l - \delta \leqq t \leqq t_0 + l \\ 0 & t \geqq t_0 + l\ ; \end{cases}$$

thus $\lambda(t)$ is continuous and $0 \leqq \lambda \leqq \chi_{J'}$.

Set $\psi = T_l^+(\lambda\,|\varphi|)$, whence $\psi \in \mathbf{F}$, $\psi \neq 0$, $|\psi|_{\mathbf{F}} \leq |\varphi|_{\mathbf{F}}$; ψ is continuous on J and vanishes outside the interval $l + J'$. Also,

$$C' = \int_{l+J'} \psi(t)\,dt \geq \int_{t_0+\delta}^{t_0+l-\delta} |\varphi(t)|\,dt \geq \int_{J'} |\varphi(t)|\,dt - 2M\delta = C(1-\varepsilon)\,.$$

We have, for $t \in 2l + J'$,

$$\int_0^{2l} \psi(t-u)\,du = \int_{t-2l}^{t} \psi(u)\,du = \int_{l+J'} \psi(u)\,du = C'\,.$$

Since ψ is continuous and vanishes outside $l + J'$, it is uniformly continuous, and there exists a positive integer n so large that $t', t'' \in J$, $|t'' - t'| \leq l/n$ implies $|\psi(t'') - \psi(t')| \leq C'\varepsilon/2l$.

We set $\omega = (l/C'n)\,\chi_{2l+J'} \sum_{i=0}^{2n-1} T_{il/n}^+\psi$; then $\omega \in \mathbf{F}$, $|\omega|_{\mathbf{F}} \leq (2l/C')\,|\psi|_{\mathbf{F}} \leq$ $\leq (2l/C')\,|\varphi|_{\mathbf{F}}$. For $t \in 2l + J'$ we have

$$|1 - \omega(t)| = (1/C')\left|\int_0^{2l} \psi(t-u)\,du - (l/n)\sum_{i=0}^{2n-1} \psi(t - il/n)\right| \leq$$

$$\leq (1/C') \sum_{i=0}^{2n-1} \int_{il/n}^{(i+1)l/n} |\psi(t-u) - \psi(t - il/n)|\,du \leq$$

$$\leq (2n/C')\,(l/n)\,(C'\varepsilon/2l) = \varepsilon\,.$$

Hence $0 \leq (1-\varepsilon)\,\chi_{2l+J'} \leq \omega$; then $\chi_{2l+J'} \in \mathbf{F}$ (so that we may take $J'' = 2l + J'$) and $|\chi_{2l+J'}|_{\mathbf{F}} \leq (1-\varepsilon)^{-1}\,|\omega|_{\mathbf{F}} \leq (1-\varepsilon)^{-1}(2l/C')\,|\varphi|_{\mathbf{F}} \leq (1-\varepsilon)^{-2}(2l/C)\,|\varphi|_{\mathbf{F}}$. Since the inequality between the extreme members holds for arbitrarily small ε, we have $C\,|\chi_{2l+J'}|_{\mathbf{F}} \leq 2l\,|\varphi|_{\mathbf{F}}$, as was to be proved.

Theorem 2. *If $\mathbf{F} \in \mathscr{T}$ is locally closed and J' is any bounded interval contained in $[\theta, \infty)$ (where $\theta = \theta(\mathbf{F})$) with $\mu(J') = l > 0$, then $\chi_{J'} \in \mathbf{F}$ and* (2) *holds, where $\alpha(l) = \alpha(\mathbf{F}; l)$.*

Proof. It is sufficient to assume $J' = [\theta + 2l, \theta + 3l]$, the statement then following in general from Corollary F. 4.2. For this purpose it is sufficient to exhibit, for every $\varepsilon > 0$, a function $\omega \in \mathbf{F}$, with $|\omega|_{\mathbf{F}} \leq (1+\varepsilon)\,(2l/\alpha(l))$, and such that $\omega(t)$ vanishes outside J' and satisfies $\int_{J'} |1 - \omega(t)|\,dt \leq \varepsilon$.

Since $[\theta, \theta + l] = -2l + J'$, there exists $\varphi \in \mathbf{F}$, $\varphi \neq 0$, such that

$$C = \int_{-2l+J'} |\varphi(t)|\,dt \geq (1+\varepsilon)^{-1}\alpha(l)\,|\varphi|_{\mathbf{F}} > 0$$

(Corollary F. 4.3). We set $\psi = T_l^+(\chi_{-2l+J'}\,|\varphi|)$, whence $\psi \in \mathbf{F}$, $|\psi|_{\mathbf{F}} \leq |\varphi|_{\mathbf{F}} \leq$ $\leq (1+\varepsilon)C/\alpha(l)$. We have, for $t \in J'$,

$$\int_0^{2l} \psi(t-u)\,du = \int_{t-2l}^{t} \psi(u)\,du = \int_{-l+J'} \psi(u)\,du = \int_{-2l+J'} |\varphi(u)|\,du = C\,.$$

Since $\psi(t)$ is integrable and vanishes essentially outside $-l + J'$, there exists a *continuous* function $\psi'(t)$, vanishing outside the same interval, and such that $\int_J |\psi(t) - \psi'(t)|\,dt \leq C\varepsilon/6l$. Since $\psi'(t)$ is uniformly continuous on J,

there exists a positive integer n so large that $t', t'' \in J$, $|t''-t'| \leqq l/n$ implies $|\psi'(t'')-\psi'(t')| \leqq C\varepsilon/6l^2$. For any $u', u'' \in J$, $|u''-u'| \leqq l/n$ we therefore have

$$(6)\quad \int_{J'} |\psi(t-u'')-\psi(t-u')|\,dt \leqq \int_J |\psi(t-u'')-\psi'(t-u'')|\,dt + \\ + \int_J |\psi(t-u')-\psi'(t-u')|\,dt + \int_{J'} |\psi'(t-u'')-\psi'(t-u')|\,dt \leqq C\varepsilon/2l\,.$$

Now set $\omega = (l/Cn)\,\chi_{J'} \sum_{i=0}^{2n-1} T^{+}_{il/n}\psi$; thus $\omega \in \mathbf{F}$, $|\omega|_{\mathbf{F}} \leqq (2l/C)\,|\psi|_{\mathbf{F}} \leqq \leqq (1+\varepsilon)\,(2l/\alpha(l))$ as required; also, $\omega(t)$ vanishes outside J'. By Fubini's theorem and by (6) we have

$$\int_{J'} |1-\omega(t)|\,dt = (1/C)\int_{J'}\left|\int_0^{2l}\psi(t-u)\,du - (l/n)\sum_{i=0}^{2n-1}\psi(t-il/n)\right| dt \leqq$$
$$\leqq (1/C)\sum_{i=0}^{2n-1}\int_{J'} dt \int_{il/n}^{(i+1)l/n} |\psi(t-u)-\psi(t-il/n)|\,du =$$
$$= (1/C)\sum_{i=0}^{2n-1}\int_{il/n}^{(i+1)l/n} du \int_{J'} |\psi(t-u)-\psi(t-il/n)|\,dt \leqq$$
$$\leqq (2n/C)\,(l/n)\,(C\varepsilon/2l) = \varepsilon$$

as claimed.

Corollary 1. *If $\mathbf{F} \in \mathscr{T}$ is quasi locally closed there exists $k > 0$ such that if J' is any bounded interval contained in $[\theta, \infty)$ with $\mu(J') = l > 0$, then $\chi_{J'} \in \mathbf{F}$ and $|\chi_{J'}|_{\mathbf{F}} \leqq kl/\alpha(l)$, where $\alpha(l) = \alpha(\mathbf{F}; l)$.*

Proof. We have $\theta(\mathbf{F}) = \theta(\mathrm{lc}\mathbf{F})$ and $\alpha(\mathbf{F}; l) = \alpha(\mathrm{lc}\mathbf{F}; l)$ for all $l > 0$. Since $\mathbf{F}$ and $\mathrm{lc}\mathbf{F}$ are norm-equivalent by assumption, there exists $\beta > 0$ such that $\mathrm{lc}\mathbf{F} \leqq \beta\mathbf{F}$. By Theorem 2, $\chi_{J'} \in \mathrm{lc}\mathbf{F}$, whence $\chi_{J'} \in \mathbf{F}$, and $|\chi_{J'}|_{\mathbf{F}} \leqq \beta\,|\chi_{J'}|_{\mathrm{lc}\mathbf{F}} \leqq \leqq 2\beta l/\alpha(\mathrm{lc}\mathbf{F}; l) = 2\beta l/\alpha(\mathbf{F}; l)$; the conclusion therefore holds with $k = 2\beta$.

Example 2. In Theorems 1 and 2 it is not possible in general to replace the constant 2 in (4) and (2), respectively, by any smaller number. Consider, in fact, the space $\mathbf{T}$, which is locally closed (Theorem F. 4.14) and belongs to $\mathscr{T}^{\#}$. Set $J' = [0, 1+\delta]$, with $0 < \delta < 1$: clearly $\chi_{J'} \in \mathbf{S}$ and, by Lemma F. 4.5, $|\chi_{J'}|_{\mathbf{T}} = |\chi_{J'}|_{\mathbf{S}} = 2$. Let φ be any continuous function such that $\chi_{[\delta/2,\,1-\delta/2]} \leqq \leqq \varphi \leqq \chi_{[0,\,1]}$ (such a function certainly exists). Then clearly $\varphi \in \mathbf{T}$, $|\varphi|_{\mathbf{T}} = 1$. But then we find

$$\frac{|\chi_{J'}|_{\mathbf{T}} \int_{J'} |\varphi(t)|\,dt}{(1+\delta)\,|\varphi|_{\mathbf{T}}} \geqq 2\,\frac{1-\delta}{1+\delta}\,,$$

and this can be made as close as we please to 2 if δ is taken sufficiently small. This at once establishes both our claims.

References

[1] SCHÄFFER, J. J.: Function spaces with translations. Math. Ann. **137**, 209—262 (1959).

(Eingegangen am 20. April 1959)

Volkmann, B.
Math. Annalen 138, 145—154 (1959)

Die Dimensionsfunktion von Punktmengen

Von

Bodo Volkmann in Mainz

1. Einleitung

Ist M eine beliebige Punktmenge im n-dimensionalen euklidischen Raum R_n, so bezeichnen wir für jeden Punkt $x \in R_n$, wenn $K_\varepsilon(x)$ die n-dimensionale Kugel um x mit dem Radius ε ist, den Grenzwert¹)

$$\lim_{\varepsilon \to 0} \dim M \cap K_\varepsilon(x) = \dim(x, M)$$

als die *Hausdorffsche Dimension von M am Punkt x*. Dabei bedeutet „dim" auf der linken Seite die gewöhnliche Hausdorffsche Dimension, die man folgendermaßen definieren kann:

Für reelles $\eta > 0$ heißt $V_\eta = \{\mathfrak{o}_1, \mathfrak{o}_2, \ldots\}$ eine η-Überdeckung von M, wenn die $\mathfrak{o}_\nu$ offene Mengen mit²) $\bigcup_{\nu=1}^{\infty} \mathfrak{o}_\nu \supseteq M$ und $|\mathfrak{o}_\nu| < \eta$ sind. Ferner heißt der Grenzwert

$$\{M\}^\alpha = \lim_{\eta \to 0} \underset{V_\eta}{\underline{\text{fin}}} \sum_{\nu=1}^{\infty} |\mathfrak{o}_\nu|^\alpha \qquad (0 < \alpha \leqq n),$$

wobei V_η alle η-Überdeckungen von M durchläuft, α-dimensionales Hausdorffsches Maß von M. Es gibt dann stets eine Zahl $\delta = \dim M$, genannt Hausdorffsche Dimension von M, bei der, soweit definiert,

$$\{M\}^\alpha = \begin{cases} \infty & \text{für} \quad \alpha < \delta, \\ 0 & \text{für} \quad \alpha > \delta \end{cases}$$

wird.

Die Funktion $f(x) = \dim(x, M)$ wollen wir auch *Dimensionsfunktion* der Menge M nennen.

In der vorliegenden Arbeit werden einige notwendige (§ 2) und einige hinreichende (§ 3) Bedingungen dafür angegeben, wann eine vorgegebene Funktion Dimensionsfunktion einer geeigneten Punktmenge ist. Dabei zeigt sich insbesondere, daß alle Dimensionsfunktionen eine Eigenschaft besitzen, die zwischen Stetigkeit und Halbstetigkeit nach oben liegt und daher von uns starke Halbstetigkeit nach oben genannt wird (Satz 1). In § 4 wird eine Invarianzeigenschaft bewiesen, die man als Analogon für Hausdorffsche

¹) Wie aus einer brieflichen Mitteilung von Herrn S. J. Taylor an den Verfasser hervorgeht, findet sich derselbe Begriff auch in der demnächst erscheinenden Arbeit [8].

²) Den Durchmesser einer Menge M bezeichnen wir mit $|M|$, ihre Ableitung mit M' und ihre abgeschlossene Hülle mit M^*.

Dimensionen zu der topologischen Invarianz der gewöhnlichen Dimension betrachten kann. Schließlich ergeben sich Anwendungen der eingeführten Begriffe auf spezielle Typen von Mengen (§§ 5 und 6).

2. Notwendige Bedingungen

Aufgrund einfacher Eigenschaften des Hausdorffschen Maßes erhält man folgende Bemerkungen als unmittelbare Folgerungen aus der obigen Definition:

1) Für jede Menge $M \subseteq R_n$ ist überall

$$0 \leqq \dim(x, M) \leqq \dim M \leqq n .$$

2) Ist $x \notin M'$, so gilt $\dim(x, M) = 0$.

3) Ist $\dim M = \delta$, so gilt

$$\overline{\operatorname{fin}}_{x \in R_n} \dim(x, M) = \overline{\operatorname{fin}}_{x \in M} \dim(x, M) = \delta .$$

Beweis. Es ist $\overline{\operatorname{fin}}_{x \in R_n} \dim(x, M) \leqq \delta$; denn andernfalls gäbe es eine Kugel $K_\varepsilon(x)$ mit $\dim M \cap K_\varepsilon(x) > \delta$, woraus $\dim M > \delta$ folgen würde. Ferner ist die Zahl $\overline{\operatorname{fin}}_{x \in M} \dim(x, M) \geqq \delta$; denn wäre sie gleich $\delta_1 < \delta$, so gäbe es zu jedem $x \in M$ ein $\varepsilon > 0$ mit $\dim M \cap K_\varepsilon(x) \leqq \delta_1$. Da dann nach dem Borelschen Überdeckungssatz eine abzählbare Teilmenge $\{K^1, K^2, \ldots\}$ aller dieser Kugeln existieren würde, die M überdeckt, so ergäbe sich der Widerspruch

$$\delta = \dim M \leqq \overline{\operatorname{fin}}_{i=1,2,\ldots} \dim M \cap K^i \leqq \delta_1 .$$

Also ist

$$\delta \geqq \overline{\operatorname{fin}}_{x \in R_n} \dim(x, M) \geqq \overline{\operatorname{fin}}_{x \in M} \dim(x, M) \geqq \delta ,$$

woraus die Behauptung folgt.

4) Ist $M_1 \subseteq M_2$, so gilt überall $\dim(x, M_1) \leqq \dim(x, M_2)$.

5) Mit der unteren bzw. oberen α-Dichte der Menge M am Punkt x, die man bekanntlich[3]), wenn v das n-dimensionale Volumen bezeichnet, als

$$\underline{D}_\alpha(x, M) = \varliminf_{\eta \to 0} \frac{\{M \cap K_\varepsilon(x)\}^\alpha}{v(K_\varepsilon(x))^\alpha} \quad \text{bzw.} \quad \overline{D}_\alpha(x, M) = \varlimsup_{\eta \to 0} \frac{\{M \cap K_\varepsilon(x)\}^\alpha}{v(K_\varepsilon(x))^\alpha}$$

definiert, steht die Dimensionsfunktion in folgendem Zusammenhang:

a) Ist $\dim(x, M) < \alpha$, so $\underline{D}_\alpha(x, M) = \overline{D}_\alpha(x, M) = \infty$.

b) Ist $\dim(x, M) > \alpha$, so $\underline{D}_\alpha(x, M) = \overline{D}_\alpha(x, M) = 0$.

c) Ist $\underline{D}_\alpha(x, M) > 0$, so $\dim(x, M) \geqq \alpha$.

d) Ist $\overline{D}_\alpha(x, M) < \infty$, so $\dim(x, M) \leqq \alpha$.

Wir nennen eine Funktion $f(x)$ ***stark halbstetig nach oben***, wenn die zugehörige Funktion[4])

$$\Phi(x) = \lim_{\varepsilon \to 0} \overline{\operatorname{fin}}_{0 < |y-x| \leqq \varepsilon} f(y) \tag{1}$$

für alle $x \in R_n$ der Gleichung $\Phi(x) = f(x)$ genügt. (Wenn man hierin die Bedingung $0 < |y - x| \leqq \varepsilon$ durch $|y - x| \leqq \varepsilon$ ersetzt, geht $\Phi(x)$ in die obere

---

[3]) Siehe A. S. BESICOVITCH [2].

[4]) Dabei bezeichnet, wie üblich, $|y - x|$ den Abstand der beiden Punkte x und y im R_n.

Limesfunktion und damit die starke Halbstetigkeit in die gewöhnliche Halbstetigkeit über.) Man erkennt unmittelbar, daß jede stetige Funktion auch stark halbstetig nach oben und jede nach oben stark halbstetige Funktion auch nach oben halbstetig ist. Dagegen zeigt schon im R_1 das Beispiel

$$f(x) = \begin{cases} 1 \text{ für } x = 1, \\ 0 \text{ für } x \neq 1, \end{cases}$$

daß eine nach oben halbstetige Funktion nicht nach oben stark halbstetig zu sein braucht, und das Beispiel

$$f(x) = \begin{cases} 1 \text{ für } x \leqq 1, \\ 0 \text{ für } x > 1, \end{cases}$$

daß nicht jede nach oben stark halbstetige Funktion stetig ist. Es gilt nun der

Satz 1. *Jede Dimensionsfunktion ist stark halbstetig nach oben.*

Beweis. Sei $f(x)$ eine Dimensionsfunktion, $\Phi(x)$ die zugehörige, in (1) definierte Funktion und M eine zugehörige Punktmenge. Dann gilt aufgrund von Bem. 3 für jeden Punkt $x \in R_n$, wenn mit $k_\varepsilon(x)$ die Menge der von x verschiedenen Punkte in der Kugel $K_\varepsilon(x)$ bezeichnet wird,

$$\Phi(x) = \lim_{\varepsilon \to 0} \dim M \cap k_\varepsilon(x) = \dim(x, M) = f(x),$$

wie behauptet.

Daß die Umkehrung von Satz 1 falsch ist, auch bei der trivialerweise notwendigen Beschränkung auf Funktionen mit Werten im Intervall $[0, n]$, ergibt sich aus dem folgenden

Satz 2. *Sei $f(x)$ eine nicht identisch verschwindende Dimensionsfunktion, σ eine Zahl mit $0 \leqq \sigma < \overline{\text{fin}}_{x \in R_n} f(x)$ und $Q(\sigma)$ die Menge der $x \in R_n$ mit $f(x) > \sigma$. Dann gilt für alle $x \in Q(\sigma)$ die Ungleichung*

$$\dim(x, Q(\sigma)) > \sigma. \tag{2}$$

Beweis. Ist M eine Menge mit der Dimensionsfunktion $f(x)$, so bezeichnen wir, wenn etwa $\dim M = \delta$ ist, mit $M(\sigma)$ die Menge der Punkte $x \in M$, für die $f(x) \leqq \sigma$ gilt, so daß $M - M(\sigma) \subseteq Q(\sigma)$ wird. Dann gilt nach Bem. 3 und 4

$$\dim M(\sigma) = \overline{\text{fin}}_{x \in M(\sigma)} \dim(x, M(\sigma)) \leqq \overline{\text{fin}}_{x \in M(\sigma)} \dim(x, M) \leqq \sigma$$

und ferner[5]) für jedes $\varepsilon > 0$ und jedes $x \in Q(\sigma)$

$$\dim(M - M(\sigma)) \cap K_\varepsilon(x) = \dim M \cap K_\varepsilon(x),$$

also

$$\dim(x, M - M(\sigma)) = \dim(x, M) = f(x) > \sigma$$

und damit auch

$$\dim(x, Q(\sigma)) > \sigma,$$

wie behauptet.

Eine Funktion, die zwar nach oben stark halbstetig ist und nur Werte in $[0, 1]$ annimmt, aber keine Dimensionsfunktion ist, erhält man mit Hilfe von Satz 2 z. B. folgendermaßen:

---

[5]) Bekanntlich ändert sich die Hausdorffsche Dimension einer Menge nicht, wenn eine Untermenge kleinerer Dimension fortgelassen wird.

Sei C das Cantorsche Diskontinuum und δ eine Zahl mit $\frac{\log 2}{\log 3} < \delta < 1$. Dann ist die Funktion

$$f(x) = \begin{cases} \delta \text{ für } x \in C, \\ 0 \text{ für } x \notin C \end{cases}$$

offenbar stark halbstetig nach oben; jedoch erfüllt sie für Zahlen σ mit $\frac{\log 2}{\log 3} < \sigma < \delta$ wegen $\dim C = \frac{\log 2}{\log 3}$ nicht die Ungleichung (2), da offenbar $\dim Q(\sigma) = \dim C < \sigma$ ist. Es existiert also keine Punktmenge mit der Dimensionsfunktion $f(x)$.

3. Hinreichende Bedingungen

Bei dem Versuch, hinreichende Bedingungen nach Art von (2) für die Eignung einer vorgegebenen Funktion als Dimensionsfunktion zu finden, stößt man auf erhebliche Schwierigkeiten, weil sich anscheinend keine naheliegende Methode anbietet, mit der man zu einer vorgegebenen Dimensionsfunktion eine Punktmenge explizit konstruieren könnte. Bei praktisch allen bisher untersuchten Typen von Punktmengen sind die zugehörigen Dimensionsfunktionen von besonders einfacher Beschaffenheit, nämlich entweder überall konstant[6]) oder — wie beim Cantorschen Diskontinuum — auf einer perfekten Menge konstant und auf der Komplementärmenge gleich Null. Dennoch gilt der

Satz 3. *Jede stetige Funktion $f(x)$ auf dem R_n mit Werten im Intervall* $[0, n]$ *ist Dimensionsfunktion.*

Beweis. Wir beschränken uns auf den Fall $n = 1$, da sich der folgende Beweis mit Hilfe von Maxfieldschen Zahlen-n-tupeln[7]) auf beliebiges $n \geqq 1$ übertragen läßt. Ferner genügt es offenbar, den Existenzbeweis für ein abgeschlossenes Intervall zu führen; wir wählen der Einfachheit halber das Intervall $[0, 1]$.

Sei $d(\delta)$ die in [10] eingeführte Funktion

$$\sigma = d(\delta) = \frac{\delta \log \delta + (1-\delta)\log(1-\delta)}{-\log 2} \qquad (0 < \delta < 1;\, d(0) = d(1) = 0)$$

und $\delta = d^{-1}(\sigma)$ ihre (nach der Festsetzung $0 \leqq \delta \leqq \frac{1}{2}$ eindeutige) Umkehrfunktion. Für jedes $\sigma \in [0, 1]$ sei ferner $D(\sigma)$ die Menge der Zahlen $x \in [0, 1]$, deren dyadische Entwicklung

$$x = \sum_{i=1}^{\infty} \frac{e_i}{2^i} \quad (e_i = 0 \text{ oder } 1;\ \text{unendlich viele } e_i \neq 0)$$

der Bedingung

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} e_i = d^{-1}(\sigma) = \delta$$

---

[6]) Vgl. dazu § 5.

[7]) J. E. Maxfield [7]. Der erwähnten brieflichen Mitteilung von Herrn S. J. Taylor zufolge enthält die Arbeit [4] den Existenzbeweis einer perfekten Menge E im Intervall $[0, 1]$, deren Dimensionsfunktion an allen Punkten von E mit einer vorgegebenen, stetigen, monoton wachsenden Funktion übereinstimmt. Diese Menge E stellt also ein nichttriviales Beispiel zu Satz 4 dar.

genügt. Nach [10], Satz 52[8]) ist somit

$$\dim D(\sigma) = d(d^{-1}(\sigma)) = \sigma .$$

Ferner sei $N(\delta)$ die Menge der $x \in [0, 1]$ mit $f(x) \geq \delta$, und schließlich sei

$$M = \bigcup_{0 \leq \sigma \leq 1} (D(\sigma) \cap N(\sigma)) .$$

Es soll nun gezeigt werden, daß M die Dimensionsfunktion $f(x)$ hat. Für jedes $x \in [0, 1]$ und $\varepsilon > 0$ ist wegen

$$M \cap [x - \varepsilon, x + \varepsilon] = \bigcup_{0 \leq \delta \leq 1} (D(\sigma) \cap N(\sigma) \cap [x - \varepsilon, x + \varepsilon]) ,$$

wenn $\min\limits_{|y - x| \leq \varepsilon} f(y) = \underline{\sigma}(\varepsilon, x)$ und $\max\limits_{|y - x| \leq \varepsilon} f(y) = \overline{\sigma}(\varepsilon, x)$ gesetzt wird,

$$\bigcup_{0 \leq \sigma \leq \underline{\sigma}(\varepsilon, x)} (D(\sigma) \cap [x-\varepsilon, x+\varepsilon]) \subseteq M \cap [x-\varepsilon, x+\varepsilon] \subseteq \bigcup_{0 \leq \sigma \leq \overline{\sigma}(\varepsilon, x)} D(\sigma) \cap [x-\varepsilon, x+\varepsilon] ,$$

also — wieder nach [10], Satz 52 —

$$\underline{\sigma}(\varepsilon, x) \leq \dim M \cap [x - \varepsilon, x + \varepsilon] \leq \overline{\sigma}(\varepsilon, x) ,$$

und daraus folgt durch den Grenzübergang $\varepsilon \to 0$ wegen der vorausgesetzten Stetigkeit von $f(x)$ die behauptete Gleichung

$$\dim (x, M) = f(x) .$$

Eine Funktion $f(x)$, die auf dem R_n definiert ist, heiße stückweise stetig, wenn es (höchstens abzählbar viele) offene Mengen $\mathfrak{O}_1, \mathfrak{O}_2, \ldots$ derart gibt, daß jeder Punkt $x \in R_n$ entweder zu genau einem $\mathfrak{O}_\nu$ gehört oder Randpunkt von mindestens einem $\mathfrak{O}_\nu$ ist und wenn außerdem $f(x)$ auf jeder Menge $\mathfrak{O}_\nu$ stetig ist. Dann gilt als Verschärfung des vorangehenden Satzes der

Satz 4. *Jede stückweise stetige, nach oben stark halbstetige Funktion $f(x)$ auf dem R_n mit Werten in $[0, n]$ ist Dimensionsfunktion.*

Beweis. Für jeden Index ν sei M_ν eine (nach Satz 3 vorhandene) Punktmenge mit $M_\nu \subseteq \mathfrak{O}_\nu$ und

$$\dim (x, M_\nu) = f(x) \quad \text{für alle} \quad x \in \mathfrak{O}_\nu \qquad (\nu = 1, 2, \ldots) .$$

Setzt man dann $M = \bigcup\limits_{\nu=1}^{\infty} M_\nu$, so gilt für alle $x \in \mathfrak{O}_\nu$

$$\dim (x, M) = \dim (x, M_\nu) = f(x) \qquad (\nu = 1, 2, \ldots) .$$

Somit stimmen die beiden Funktionen $f(x)$ und $\dim (x, M)$ auf der Menge $\bigcup\limits_{\nu=1}^{\infty} \mathfrak{O}_\nu$ überein, also — da beide stark halbstetig nach oben sind — auch auf deren abgeschlossener Hülle, d. h. überall im R_n.

4. Eine Invarianzeigenschaft

Eine eineindeutige Abbildung T einer offenen Menge $\mathfrak{O}_1 \subseteq R_n$ auf eine offene Menge $\mathfrak{O}_2 \subseteq R_n$ heiße *lokalbeschränkt*, wenn es zu jedem Punkt $x \in \mathfrak{O}_1$

---

[8]) Ähnliche Aussagen wurden zuerst von BESICOVITCH [1] und KNICHAL [5] bewiesen.

positive Zahlen $c_1(x)$, $c_2(x)$ und $c_3(x)$ mit $c_1(x) \leqq c_2(x)$ so gibt, daß für alle Punktepaare x', x'' mit $|x - x'| \leqq c_3(x)$, $|x - x''| \leqq c_3(x)$ die Ungleichungen

$$c_1(x) \leqq \frac{|T(x') - T(x'')|}{|x' - x''|} \leqq c_2(x)$$

erfüllt sind. Dies ist offenbar äquivalent mit der Bedingung, daß die Abbildung T auf jeder kompakten Teilmenge von $\mathfrak{O}_1$ beschränkt sein soll. Wie man leicht verifiziert, bildet die Menge aller lokalbeschränkten Abbildungen von $\mathfrak{O}_1$ auf $\mathfrak{O}_2$ eine Gruppe $\mathfrak{L}(\mathfrak{O}_1, \mathfrak{O}_2)$, die in der Gruppe $\mathfrak{T}(\mathfrak{O}_1, \mathfrak{O}_2)$ aller topologischen Abbildungen von $\mathfrak{O}_1$ auf $\mathfrak{O}_2$ enthalten ist. Es gilt der

Satz 5. *Ist $T \in \mathfrak{L}(\mathfrak{O}_1, \mathfrak{O}_2)$, so ist für jede Menge $M \subseteq \mathfrak{O}_1$ und jeden Punkt $x \in \mathfrak{O}_1$*

$$\dim(T(x), T(M)) = \dim(x, M)\,.$$

Beweis. Sei $x \in \mathfrak{O}_1$, $0 < \varepsilon < \frac{1}{2} c_3(x)$, $0 < \eta < \frac{\varepsilon}{2}$ und V_η eine η-Überdeckung von $M \cap K_\varepsilon(x)$ im Sinne von § 1. Außerdem soll V_η noch so gewählt sein, daß alle als Elemente auftretenden offenen Mengen $\mathfrak{o}_\nu$ in der Kugel $K_{c_3(x)/2}(x)$ enthalten sind. Dann geht V_η durch die Abbildung T in ein System $\{T(\mathfrak{o}_1), T(\mathfrak{o}_2), \ldots\}$ von Mengen über, die wegen der vorausgesetzten Lokalbeschränktheit offen sind und ferner der Bedingung

$$|T(\mathfrak{o}_\nu)| \leqq c_2(x)\,|\mathfrak{o}_\nu| \leqq c_2(x)\,\eta$$

genügen. Da die Bildmenge $T(K_\varepsilon(x))$ offen ist und sicher wegen $T(x) \in T(K_\varepsilon(x))$ eine Kugel $K_\gamma(T(x))$ mit

$$0 < \gamma \leqq |T(K_\varepsilon(x))| \leqq c_2(x)\varepsilon \tag{3}$$

enthält, ist somit das Mengensystem $T(V_\eta) = \{T(\mathfrak{o}_1), T(\mathfrak{o}_2), \ldots\}$ eine $c_2(x)\,\eta$-Überdeckung von $K_\gamma(T(x)) \cap T(M)$ mit

$$\sum_{\nu=1}^{\infty} |T(\mathfrak{o}_\nu)|^\alpha \leqq c_2(x)^\alpha \sum_{\nu=1}^{\infty} |\mathfrak{o}_\nu|^\alpha \qquad (0 < \alpha \leqq n)\,. \tag{4}$$

Wendet man diese Überlegung auf alle η-Überdeckungen V_η von $M \cap K_\varepsilon(x)$ an und bildet man auf beiden Seiten von (4) die untere Grenze, so ergibt sich schließlich durch den Grenzübergang $\varepsilon \to 0$, der wegen (3) den Grenzübergang $\gamma \to 0$ nach sich zieht, die Ungleichung

$$\dim(T(x), T(M)) \leqq \dim(x, M)\,.$$

Da man durch Betrachtung der inversen Abbildung T^{-1} auch die entgegengesetzte Abschätzung erhält, ergibt sich die Behauptung.

Folgerung. *Die Hausdorffsche Dimension einer Punktmenge ist gegenüber lokalbeschränkten Abbildungen invariant.*

5. Dimensionshomogene Mengen

Ist M eine Untermenge eines euklidischen Raumes R und $\mathfrak{O} \in R$ offen, so nennen wir M in $\mathfrak{O}$ *dimensionshomogen*, wenn ihre Dimensionsfunktion dort konstant ist. Dieser Begriff ist verwandt mit dem von Knopp [6] in Anlehnung

an VAN VLECK [9] eingeführten Begriff der Homogenität einer linearen Punktmenge M, bei dem gefordert wird, daß der Quotient

$$\frac{\mu(M\cap[x-\varepsilon, x+\varepsilon])}{2\varepsilon} \quad (\mu = \text{äußeres Lebesguesches Maß})$$

in jedem Intervall $[x-\varepsilon, x+\varepsilon]$ denselben Wert annimmt. Bekanntlich[9]) kann dieser Wert dann nur Null oder Eins sein, so daß eine Menge genau dann homogen ist, wenn entweder sie selbst oder ihre Komplementärmenge fast alle Punkte enthält. Somit ist eine dimensionshomogene Menge M im R_n jedenfalls auch homogen, falls $\dim M < n$ ist. Jedoch geht der Zusammenhang zwischen beiden Begriffen wesentlich weiter, da sich die folgende, von KNOPP[10]) angegebene hinreichende Bedingung für die Homogenität von linearen Punktmengen auf Dimensionshomogenität übertragen läßt:

Sei M eine meßbare lineare Menge, zu der bei jedem reellen x entweder alle Zahlen $\{2^m x\} = 2^m x - [2^m x]$ ($m \geqq 0$, ganz) oder keine von ihnen gehören. Dann ist M homogen.

Als Verschärfung dieses Resultates beweisen wir den

Satz 6. *Sei $g \geqq 2$ eine ganze Zahl und $M \subseteq R_n$ eine Menge, zu der bei jedem $x \in R_n$ entweder alle Punkte[11]) $\{g^m x\}$ oder keine von ihnen gehören. Dann ist M dimensionshomogen.*

Beweis. Ist M eine Menge, die den Voraussetzungen des Satzes genügt, so betrachten wir alle Punkte der Form

$$x' = T_{m\,k}(x) = \left\{x + \frac{k}{g^m}\right\}, \quad x \in M,$$

wobei m eine ganze Zahl und $k = (k_1, \ldots, k_n)$ ein Gitterpunkt ist. Dann ist wegen $x \in M$ nach Voraussetzung $\{g^m x\} \in M$, also wegen $\{g^m x'\} = \{g^m x\}$ auch $x' \in M$, so daß die Menge M bei jeder Translation $T_{m\,k}$ in sich übergeht. Daher gilt, wenn etwa $\dim M = \delta$ und x ein Punkt in M mit $\dim(x, M) \geqq \delta - \varepsilon$ $(0 < \varepsilon < \delta)$ ist, für alle m und k

$$\dim(x, M) = \dim(x', T_{m\,k}(M)) = \dim(x', M) \geqq \delta - \varepsilon\,.$$

Da die Bildpunkte x' überall dicht liegen, gilt wegen der starken Halbstetigkeit der Dimensionsfunktion diese Ungleichung für alle $x \in R_n$, und somit ist, wie behauptet, im ganzen Raum

$$\dim(x, M) = \delta\,.$$

Eine weitere Klasse von dimensionshomogenen Mengen erhält man mit Hilfe von

Satz 7. *Eine Punktmenge $G \subseteq R_n$ ist dimensionshomogen in dem kleinsten sie enthaltenden linearen Teilraum, wenn sie bezüglich der Vektoraddition eine Gruppe bildet.*

Beweis. Sei G eine solche additive Gruppe. Ist dann der Nullpunkt kein Häufungspunkt von G, so ist die Gruppe abzählbar und folglich überall

[9]) Siehe [6].

[10]) [6], Satz 6.

[11]) Dabei wird, wenn $x = (x_1, \ldots, x_n)$ ist, $\{g^m x\} = (\{g^m x_1\}, \ldots, \{g^m x_n\})$ gesetzt.

$\dim(x, G) = 0$. Andernfalls erkennt man leicht, daß außer dem Nullpunkt auch jeder andere Punkt von G ein Häufungspunkt, also G in sich dicht ist. Dann folgt aus der Gruppeneigenschaft, daß die abgeschlossene Hülle G^* jede Gerade durch den Nullpunkt enthält, auf der unendlich viele Punkte von G liegen, und daß G^* folglich mit dem von G erzeugten linearen Raum R übereinstimmt. Ist nun $\eta > 0$ und $x \in R$ ein Punkt mit

$$\dim(x, G) > \dim G - \eta\,,$$

so wird für alle Punkte $x + y$ mit $y \in G$ offenbar

$$\dim(x + y, G) = \dim(x, G) > \dim G - \eta\,,$$

da für jedes $\varepsilon > 0$ die Menge $K_\varepsilon(x) \cap G$ bei der Translation $x' = x + y$ in die Menge $K_\varepsilon(x + y) \cap G$ übergeht. Somit gilt überall in G, also wegen der starken Halbstetigkeit von $\dim(x, G)$ auch überall in R die Ungleichung

$$\dim G \geqq \dim(x, G) > \dim G - \eta\,,$$

woraus sich durch den Grenzübergang $\eta \to 0$ die Behauptung ergibt.

Weiter beweisen wir für lineare Mengen den

Satz 8. *Eine multiplikative Gruppe G reeller Zahlen ist, wenn sie nur positive Elemente hat, in der positiven Halbgeraden R_1^+, sonst im ganzen R_1 dimensionshomogen.*

Beweis. a) Besteht G nur aus positiven Zahlen, so erhält man bei der Abbildung $T(x) = \log x$ von R_1^+ auf R_1 als Bildmenge $T(G)$ eine additive Gruppe, die nach Satz 7 dimensionshomogen ist. Da die Abbildung offenbar lokalbeschränkt ist, folgt somit die Behauptung nach Satz 5.

b) Enthält G ein negatives Element x_0 und bezeichnet man mit G^+ die Untergruppe aller positiven sowie mit G^- die Untermenge aller negativen Elemente von G, so ist $G^- = x_0 G^+$. Somit entsteht G^- aus der Gruppe G^+ durch eine lokalbeschränkte Abbildung. Da G^+ nach Teil a) in R_1^+ dimensionshomogen ist, trifft dies also nach Satz 5 auch für G^- und somit wegen[12]) $\dim G^+ = \dim G^-$ auch für G zu.

6. Dimensionslose Mengen

Manche Autoren[13]) bezeichnen eine beschränkte Punktmenge M als *dimensionslos*, wenn das Maß $\{M\}^{\dim M}$ entweder Null oder Unendlich ist. Eine beliebige Menge M soll dimensionslos heißen, wenn jede beschränkte Teilmenge von M diese Eigenschaft hat. Da bisher, soweit dem Verfasser bekannt, nur einzelne Beispiele für solche Mengen in der Literatur angegeben wurden, dürften die beiden folgenden Sätze von Interesse sein, von denen jeder die Existenz einer ausgedehnten Klasse von dimensionslosen Mengen zur Folge hat.

Satz 9. *Jede Menge $M \subseteq R_n$, die für ein $g \geqq 2$ der Voraussetzung von Satz 6 und der Bedingung $\dim M < n$ genügt*[14]), *ist dimensionslos.*

[12]) Vgl. die Folgerung zu Satz 5.

[13]) Siehe z. B. E. Best [5].

[14]) Dies ist bei zahlreichen Typen von Mengen der Fall, die durch asymptotische Ziffereneigenschaft charakterisiert sind; siehe z. B. [11], [12], [13].

Beweis. Ist M eine solche Menge, so genügt, da M durch Verschiebung um Gittervektoren in sich übergeht, der Nachweis der Behauptung für den Durchschnitt von M mit dem Einheitswürfel[15]) E_n. Ist dann W_j $(1 \leq j \leq g^n)$ einer der g^n Würfel der Form

$$\frac{k_i}{g} \leq x_i < \frac{k_i + 1}{g} \quad (i = 1, \ldots, n; 0 \leq k_i < g; k_i \text{ ganz}),$$

in die sich E_n zerlegen läßt, so erkennt man leicht aufgrund der Voraussetzung, daß die Menge $M \cap W_j$ mit der Menge $M \cap E_n$ im Verhältnis $1 : g$ geometrisch ähnlich ist. Daher ist auch jede Überdeckung von $M \cap E_n$ durch ein abzählbares System offener Mengen mit einer solchen von $M \cap W_j$ geometrisch ähnlich und umgekehrt, so daß sich, wenn $\dim M = \delta$ gesetzt wird, die Gleichung

$$g^{-\delta} \{M \cap E_n\}^\delta = \{M \cap W_j\}^\delta$$

ergibt. Daraus folgt durch Summation über $j = 1, \ldots, g^n$

$$g^{n-\delta} \{M \cap E_n\}^\delta = \left\{M \cap \left(\bigcup_{j=1}^{g^n} W_j\right)\right\}^\delta = \{M \cap E_n\}^\delta,$$

und wegen $n - \delta \neq 0$ ist dies ein Widerspruch, sofern nicht entweder

$$\{M \cap E_n\}^\delta = 0 \text{ oder } \{M \cap E_n\}^\delta = \infty \text{ gilt.}$$

Bezeichnet man bei einer gegebenen Menge M die Menge aller $x \in M$ mit $\dim(x, M) = \dim M$ als P_M, so gilt der

Satz 10. *Sei $M \subseteq R_n$ eine Menge, die der Bedingung*[16]) $\dim P_M < \dim M$ *genügt. Dann ist M dimensionslos.*

Beweis. Sei $M_i = M\left(\delta - \frac{1}{i}\right)$ in der Bezeichnungsweise von Satz 2 $(\delta = \dim M)$. Dann ist offensichtlich

$$M = \left(\bigcup_{i=1}^{\infty} M_i\right) \cup P_M$$

und daher

$$\{M\}^\delta \leq \sum_{i=1}^{\infty} \{M_i\}^\delta + \{P_M\}^\delta. \tag{5}$$

Es gilt aber für jeden Punkt $x \in M_i$

$$\dim(x, M_i) \leq \dim(x, M) \leq \delta - \frac{1}{i},$$

also $\dim M_i \leq \delta - \frac{1}{i}$ und folglich

$$\{M_i\}^\delta = 0 \; (i = 1, 2, \ldots);$$

ferner ist nach Voraussetzung $\{P_M\}^\delta = 0$, also wegen (5) auch $\{M\}^\delta = 0$. Somit ist M dimensionslos.

---

[15]) Definiert durch die Ungleichungen $0 \leq x_i < 1$ $(i = 1, \ldots, n)$.

[16]) Diese Voraussetzung ist z. B. dann erfüllt, wenn die Dimensionsfunktion stückweise stetig ist und ihr Maximum höchstens an abzählbar vielen Stellen annimmt (vgl. Satz 4).

7. Schlußbemerkung

Bekanntlich[17]) kann man den Begriff der Hausdorffschen Dimension einer Punktmenge auf beliebige separable metrische Räume übertragen und somit auch dort die Dimension einer Menge an einem Punkt definieren. Damit lassen sich unsere Sätze 1 und 2 einschließlich der Beweise wörtlich auf Punktmengen in separablen metrischen Räumen verallgemeinern. Dasselbe gilt auch von dem Begriff der Lokalbeschränktheit einer umkehrbar eindeutigen Abbildung einer offenen Menge auf eine andere und dem damit zusammenhängenden Satz 5, ferner von dem Begriff der Dimensionslosigkeit und Satz 10. Dagegen machen die Beweise der übrigen Sätze von speziellen Eigenschaften der reellen Zahlen Gebrauch, so daß hier eine unmittelbare Verallgemeinerung auf metrische Räume nicht möglich ist.

Literatur

[1] Besicovitch, A. S.: On the sum of digits of real numbers represented in the dyadic number system. Math. Ann. **110**, 321—330 (1934). — [2] Besicovitch, A. S.: On linear sets of fractional dimensions. I. Math. Ann. **101**, 169—193 (1929). — [3] Best, E.: A closed dimensionless linear set. Proc. Edinburgh Math. Soc. **2**, (6), 105—108 (1939). — [4] Eggleston, H. G.: A characteristic property of Hausdorff measure. J. London Math. Soc. **25**, 39—46 (1950). — [5] Knichal, V.: Dyadische Entwicklungen und Hausdorffsches Maß. Mém. Soc. roy. Sci. Bohême Nr. 14, 1—18 (1933). — [6] Knopp, K.: Mengentheoretische Behandlung einiger Probleme der diophantischen Approximationen und der transfiniten Wahrscheinlichkeiten. Math. Ann. **95**, 409—420 (1926). — [7] Maxfield, J. E.: Normal k-tuples. Pacif. J. Math. **3**, 189—196 (1953). — [8] Rogers, C. A., and S. J. Taylor: The analysis of additive set functions in Euclidean space. Erscheint in Acta math. — [9] Vleck, E. B. van: On non-measurable sets of points, with an example. Trans. Amer. Math. Soc. **9**, 237—244 (1908). — [10] Volkmann, B.: Über Klassen von Mengen natürlicher Zahlen. J. reine angew. Math. **190**, 199—230 (1952). — [11] Volkmann, B.: Über Hausdorffsche Dimensionen von Mengen, die durch Zifferneigenschaften charakterisiert sind. II. Math. Z. **59**, 247—254 (1953). — [12] Volkmann, B.: Über Hausdorffsche Dimensionen von Mengen, die durch Zifferneigenschaften charakterisiert sind. IV. Math. Z. **59**, 425—433 (1954). — [13] Volkmann, B.: Über Hausdorffsche Dimensionen von Mengen, die durch Zifferneigenschaften charakterisiert sind. VI. Math. Z. **68**, 439—449 (1958).

(Eingegangen am 12. März 1959)

[17]) Siehe z. B. H. G. Eggleston [4].

CORDES, H. O.
Math. Annalen 138, 155—178 (1959)

Vereinfachter Beweis der Existenz einer Apriori-Hölderkonstanten

Von

H. O. CORDES aus Los Angeles

In einer Arbeit des Verfassers [1] konnte gezeigt werden, daß es möglich ist, für die Lösungen u eines Randwertproblems der Form

$$L(u) = \sum_{i,k=1}^{n} a_{ik}(x)\, \partial^2 u/\partial x_i \partial x_k + \sum_{i=1}^{n} b_i(x)\, \partial u/\partial x_i + b_0(x) u = f\,, \qquad x \in B$$

$$u = \varphi\,, \quad x \in \Gamma$$

eine Apriori-Hölderkonstante anzugeben, die nicht vom Stetigkeitsverhalten der Koeffizienten $a_{ik}(x)$ abhängt, falls man fordert, daß die $a_{ik}(x)$ einer sogenannten K_ε-Bedingung mit einem von x unabhängigen ε genügen. Diese K_ε-Bedingung bedeutet dabei eine Verschärfung der Elliptizitätsforderung in folgendem Sinne: Man denke sich der symmetrischen Matrix $A = ((a_{ik}))$ den Vektor

$$\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_n)$$

zugeordnet, in welchem die Komponenten $\lambda_1, \lambda_2, \ldots, \lambda_n$ die n Eigenwerte von A, in irgendeinem, wohldefinierten Sinne angeordnet, darstellen mögen. Auf diese Weise ergibt sich eine Abbildung des Raumes aller symmetrischen Matrizen auf den n-dimensionalen Raum der Vektoren $(\lambda_1, \lambda_2, \ldots, \lambda_n)$. Speziell entsprechen dabei die positiv definiten Matrizen dem (offenen), verallgemeinerten „ersten Oktanten": $\lambda_1 > 0, \lambda_2 > 0, \ldots, \lambda_n > 0$. Die Forderung, der Operator L sei gleichmäßig elliptisch, kann dann folgendermaßen ausgesprochen werden: Es gibt einen n-dimensionalen Würfel W: $p \leq \lambda_j \leq P$, $j = 1, \ldots, n$, (mit $p > 0$), derart, daß $\lambda(x) \in W$ für jedes $x \in B$. Wir betrachten nun denjenigen Kreiskegel, der als Achse die Gerade $\lambda_1 = \lambda_2 = \ldots = \lambda_n$ besitzt und der die Wände des ersten Oktanten, d. h. die Hyperebenen $\lambda_j = 0$ gerade berührt. Seine Gleichung kann geschrieben werden in der Form

$$(n-1) \sum_{i<k} (\lambda_i - \lambda_k)^2 = \left(\sum_{i=1}^{n} \lambda_i \right)^2.$$

Das Innere dieses Kegels ist ein Teil des ersten Oktanten. Es werde mit K_0 bezeichnet. Wenn $\lambda(x) \in K_0$ für alle $x \in B$, so sagen wir, der Operator L genüge einer Kegelbedingung, genauer einer K_0-Bedingung. Wenn a fortiori für den Vektor $\lambda(x)$ gilt

$$(n-1) \sum_{i<k} (\lambda_i - \lambda_k)^2 \leq (1-\varepsilon) \left(\sum_{i=1}^{n} \lambda_i \right)^2$$

mit einem $\varepsilon > 0$ unabhängig von x, so sagen wir, L genüge einer K_ε-Bedingung. Dies bedeutet also nichts anderes, als daß es einen etwas kleineren Kegel K_ε gibt, der mitsamt seinem Rande ganz im Inneren von K_0 liegt und der mit K_0 die Achse gemeinsam hat, welcher die gesamte Punktmenge $\Lambda = \{\lambda(x) \ni x \in B\}$ enthält.

Es ist offenbar leicht möglich, die oben definierte K_ε-Bedingung auch durch die Koeffizienten a_{ik} der Matrix A direkt auszudrücken. Eine leichte Rechnung zeigt, daß

$$\sum_{i<k} (\lambda_i - \lambda_k)^2 = (n-1)\left(\sum_{i=1}^{n} a_{ii}\right)^2 - n \sum_{i,k=1}^{n} (a_{ii}a_{kk} - a_{ik}^2)$$

und

$$\left(\sum_{i=1}^{n} \lambda_i\right)^2 = \left(\sum_{i=1}^{n} a_{ii}\right)^2$$

gilt; setzt man dies in die K_ε-Bedingung ein, so ergibt sich

$$n(n-1) \sum_{i,k=1}^{n} a_{ik}^2 \leq (n-\varepsilon)\left(\sum_{i=1}^{n} a_{ii}\right)^2$$

als eine zur K_ε-Bedingung äquivalente Forderung.

Ferner kann man leicht eine schärfere Bedingung angeben, die die bei der Definition der gleichmäßigen Elliptizität übliche Form besitzt:

$$m \sum_{i=1}^{n} \xi_i^2 \leq \sum_{i,k=1}^{n} a_{ik}\xi_i\xi_k \leq M \sum_{i=1}^{n} \xi_i^2$$

und die die K_ε-Bedingung garantiert. Man hat lediglich m und M so zu wählen, daß der durch die obige Bedingung charakterisierte Würfel im Raum der Vektoren $(\lambda_1, \ldots, \lambda_n)$ noch ganz innerhalb eines passenden Kegels K_ε enthalten ist.

Auch sei bemerkt, daß keineswegs etwa jeder Operator L, für den der Vektor $\lambda(x)$ nicht stets im Kegel K_0 enthalten ist, grundsätzlich von der Behandlung durch die hier diskutierten Methoden ausgeschlossen ist. Durch eine lineare Transformation der n Veränderlichen $x_1, \ldots, x_n$ kann bekanntlich stets erreicht werden, daß die Matrix $A(x)$ für ein beliebig vorgegebenes x^0 in die Einheitsmatrix übergeht, sofern nur der Operator L elliptisch, d. h. die Matrix $A(x)$ für jedes x positiv definit ist. Da der zur Einheitsmatrix gehörige Vektor λ auf der Achse des Kegels K_0 liegt, wird also nach Ausübung dieser Transformation die Kegelbedingung K_ε zumindest für den Punkt x^0 erfüllt sein. Die K_ε-Bedingung fordert daher lediglich, daß auch für alle anderen x des Bereiches B, oder wenigstens einer Umgebung von x^0, der Vektor $\lambda(x)$ sich nicht zu weit vom Vektor $\lambda(x^0)$ entferne, wobei aber Unstetigkeiten beliebiger Art zugelassen sein können.

Endlich sei angemerkt, daß speziell für $n = 2$ eine K_ε-Bedingung im wesentlichen mit der Forderung der gleichmäßigen Elliptizität äquivalent ist. Unsere Theorie liefert daher dann das volle, bereits früher von Leray-Schauder [2] und Nirenberg [4] bewiesene Resultat.

Obwohl es wünschenswert wäre, sich von der stärkeren K_ε-Bedingung zu befreien und statt dessen die Bedingung der gleichmäßigen Elliptizität zu verlangen, ist doch bisher diese Frage für den Fall $n > 2$ im allgemeinen offen geblieben. Da zudem der oben zitierte Beweis des Verfassers den Nachteil hat, recht kompliziert zu sein, und da auch kein anderer Beweis bisher bekannt geworden zu sein scheint[1]), mag es erlaubt sein, im folgenden einen zweiten, wesentlich vereinfachten, Beweis darzulegen. Bekanntlich ergibt sich aus der genannten Apriori-Abschätzung eine Existenzaussage für die Lösung des Dirichletschen Problems der quasilinearen elliptischen Differentialgleichung zweiter Ordnung.

Wir fassen das Resultat folgendermaßen zusammen:

Die Funktionen $a_{ik}(x) = a_{ik}(x_1, \ldots, x_n)$, $i, k, = 1, \ldots, n$, seien in einem beschränkten Bereich B des $(x_1, \ldots, x_n)$-Raumes mit der Berandung Γ erklärt und meßbar; es gelte $a_{ik}(x) = a_{ki}(x)$ für $i, k = 1, \ldots, n$ und ferner $p \leq \sum_{i=1}^{n} a_{ii}(x) \leq P$ mit zwei positiven Konstanten p und P für alle x aus B. Es sei $A(x) = ((a_{ik}(x)))$ gesetzt.

Zu jedem x^0 aus B gebe es eine Kugel $\mathfrak{K}_{x^0}$: $|x - x^0| \leq \tau$ und eine nichtsinguläre Matrix $T_{x^0} = ((t_{ik}^{x^0}))$ mit $q \sum_{i=1}^{n} \xi_i^2 \leq \sum_{i=1}^{n} \left| \sum_{k=1}^{n} t_{ik}^{x^0} \xi_k \right|^2 \leq Q \sum_{i=1}^{n} \xi_i^2$, q und Q positiv und unabhängig von x^0, so daß in $B = B_{x^0} \cap \mathfrak{K}_{x^0}$ die Matrix

$$A^{x^0}(x) = T'_{x^0} A(x)\, T_{x^0}$$

einer K_ε-Bedingung genügt. ε und τ seien dabei unabhängig von x und x^0.

Die Berandung Γ des Bereiches B bestehe aus endlich vielen, zweimal stetig differenzierbaren Hyperflächen, die sich weder berühren noch schneiden.

Die Funktionen $b_i(x)$, $i = 0, 1, \ldots, n$, seien in B meßbar und beschränkt, es gelte

$$\sum_{i=0}^{n} b_i^2 \leq M\,, \quad x \in B\,.$$

Die Funktion $u(x)$ sei in B stetig differenzierbar und habe stückweise stetige zweite Ableitungen[2]). Es gelte

$$L(u) = \sum_{i,k=1}^{n} a_{ik}(x)\, \partial^2 u/\partial x_i\, \partial x_k + \sum_{i=1}^{n} b_i(x)\, \partial u/\partial x_i + b_0(x) u = f\,, \quad x \in B\,,$$

$$u(x) = \varphi(x)\,, \quad x \in \Gamma\,.$$

Es sei

$$H_\alpha(u, B) = \sup_{\substack{x^1 \neq x^2 \\ x^1, x^2 \in B}} |u(x^1) - u(x^2)|\, |x^1 - x^2|^{-\alpha}\,,$$

[1]) Wie mir Herr Nirenberg mitteilte, ist er in der Lage, die in [1] hergeleiteten Apriori-Abschätzungen zu beweisen unter der Voraussetzung, daß $n(n-1) \sum_{2<k} (\lambda_i - \lambda_k)^2 \leq (1-\varepsilon)(\Sigma\, \lambda_{ii})^2$ gilt. Diese Forderung ist schärfer als die in [1] benutzte K'_ε-Bedingung, insbesondere wird sie wesentlich schärfer, wenn n sehr groß wird.

[2]) Dies bedeute, daß nach Einteilung von B in endlich viele Unterbereiche mit stückweise glatten Rändern die zweiten Ableitungen von $u(x)$ in jedem der Unterbereiche stetig sind.

ferner

$$\|f\| = \sup_{x^0 \in B} |f(x^0)| \,.$$

Endlich sei $\|\varphi\|_2'$ die C^2-Norm der Funktion φ auf dem Rande Γ, etwa im Sinne der Definition aus [1], Seite 291.

Dann gilt der

Satz 1: Es gibt ein $\alpha > 0$ und ferner Apriori-Konstanten c_1, c_2, c_3, die nach Festlegung des Bereiches B und geeigneter Zahlen $n, p, P, q, Q, \varepsilon, \tau$ und α ohne irgendeine weitere Information über die Koeffizienten $a_{ik}(x)$, $b_i(x)$ und die Funktion $u(x)$ bereits bestimmt werden können, so daß die Abschätzung

$$H_\alpha(u; B) \leqq c_1 \|f\| + c_2 \|u\| + c_3 \|\varphi\|_2'$$

gilt.

Aus obigem Satz folgt auf Grund von wohlbekannten Überlegungen der folgende Existenzsatz:

Satz 2: Die Funktionen $A_{ik}(x, u) = A_{ik}(x_1, \ldots, x_n, u)$ seien für alle Werte der Veränderlichen u und für $x \in B$ erklärt und stetig. B und seine Berandung Γ möge den oben erwähnten Voraussetzungen genügen. In jedem kompakten Teilbereich $\tilde{B}^0$ des topologischen Produktes $B^0 = \{x, u \ni x \in B, -\infty < u < +\infty\}$ möge eine gleichmäßige Hölderbedingung für die A_{ik} erfüllt sein:

$$|A_{ik}(x^1, u^1) - A_{ik}(x^2, u^2)| \leqq C[|x^1 - x^2|^2 + |u^1 - u^2|^2]$$
$$(x^1, u^1) \in \tilde{B}^0 \,, \quad (x^2, u^2) \in \tilde{B}^0 \,.$$

Die Matrix $\big((A_{ik}(x, u))\big) = A(x, u)$ sei symmetrisch, und es sei $\sum_{i=1}^{3} A_{ii} \neq 0$. Es gebe positive Konstanten τ, ε, q und Q und zu jedem $x^0 \in B$ eine nichtsinguläre Matrix $T_{x^0} = ((t_{ik}^{x^0}))$ mit $q \sum_{i=1}^{n} \xi_i^2 \leqq \sum_{i=1}^{n} \left| \sum_{k=1}^{n} t_{ik}^{x^0} \xi_k \right|^2 \leqq Q \sum_{k=1}^{n} \xi_i^2$, derart, daß für alle x aus $B_{x^0} = \{x \ni x \in B, |x - x^0| < \tau\}$ und alle u aus $-\infty < u < +\infty$ die Matrix $A^{x^0}(x, u) = T'_{x^0} A(x, u) T_{x^0}$ einer K_ε-Bedingung genügt.

Dann besitzt die quasilineare elliptische Differentialgleichung zweiter Ordnung

$$\sum_{i,k=1}^{n} A_{ik}(x, u)\, \partial^2 u / \partial x_i\, \partial x_k = 0$$

eine Lösung u, die in ganz B definiert und zweimal stetig differenzierbar ist, welche der Randbedingung

$$u(x) = \varphi(x) \,, \quad x \in \Gamma$$

genügt. $\varphi(x)$ kann dabei als eine auf Γ zweimal stetig differenzierbare Funktion willkürlich vorgegeben werden.

Der im folgenden dargelegte Beweis des Satzes 1 benutzt wesentlich wiederum jenen von C. B. MORREY [3], L. NIRENBERG [4] und K. O. FRIEDRICHS in verschiedenen Variationen begangenen Umweg über die quadratischen Integralabschätzungen. Zunächst wird benutzt, daß gleichmäßige Hölderstetigkeit im wesentlichen äquivalent ist zur Forderung der gleichmäßigen

Beschränktheit des Integrals

$$\int_B u^2 |x - x^0|^{-n-2\alpha}\, dx$$

für alle x^0 des betrachteten Bereiches B (§ 1). Unter Verwendung des Wirtingerschen Lemmas kann das obige Integral leicht durch ein ähnliches Integral über die ersten Ableitungen abgeschätzt werden, in welchem nur die Potenz von $|x - x^0|$ um zwei Einheiten erhöht ist; auf ähnliche Weise gelingt in § 3 die weitere Abschätzung durch ein Integral über die zweiten Ableitungen, in dem nunmehr die Potenz $|x - x^0|^{4-n-2\alpha}$ auftritt. Endlich wird die K_ε-Bedingung benutzt, um letzteres Integral durch ein Integral über $(L(u))^2$ und ein weiteres über u^2 abzuschätzen (§ 2), wobei nunmehr die Potenz von $|x - x^0|$ so beschaffen ist, daß man weiter durch einen Ausdruck der Form $C_1 \|L(u)\|^2 + C_2 \|u\|^2$ abschätzen kann, womit im Prinzip der Satz 1 und auch der Satz 2 bewiesen ist. Schwierigkeiten macht nur die Behandlung des Randes, die ganz wie in [1] durchgeführt und daher leider immer noch mühsam ist.

Es ist möglich, auch die Abschätzung der ersten Ableitungen und ihrer Hölderkonstanten, die ebenfalls in [1] durchgeführt wurde, ebenfalls in dem hier dargelegten Sinne zu vereinfachen, wenn man als weiteres Hilfsmittel die in [1], § 2—3, hergeleiteten Integralidentitäten heranzieht. Aus Gründen der Durchsichtigkeit der Darstellung sei jedoch hier darauf verzichtet.

§ 1. Hölderstetigkeit und Beschränktheit von $\int u^2 r^{-n-2\alpha}\, dx$

Die Funktion $u(x)$ sei in dem beschränkten, n-dimensionalen Bereich B erklärt. Wir definieren etwas abweichend vom üblichen Sprachgebrauch als δ-Hölderkonstante $H_{\alpha,\delta} = H_{\alpha,\delta}(u; B)$ der Funktion $u(x)$ im Bereich B und zum Exponenten α:

$$H_{\alpha,\delta} = \sup_{\substack{x^1 \neq x^2;\, x^1, x^2 \in B \\ |x^1 - x^2| \leq \delta}} \{|u(x^1) - u(x^2)|\, |x^1 - x^2|^{-\alpha}\}.$$

Die Konstante $\delta > 0$ sei dabei fest vorgegeben.

Wir wollen im folgenden voraussetzen, daß der Bereich B die oben in der Einleitung angegebenen Voraussetzungen erfüllt. Offenbar ist dann die Forderung der gleichmäßigen Hölderstetigkeit von $u(x)$ im Bereich B gleichbedeutend mit der Bedingung $H_{\alpha,\delta}(u; B) < \infty$. Allerdings wird im allgemeinen $H_{\alpha,\delta}(u; B)$ einen kleineren Wert haben als die früher definierte, allgemein übliche Hölderkonstante $H_\alpha(u; B)$, jedoch läßt sich leicht eine nur von α, δ, n und der Art des Bereiches B abhängende Zahl c finden, derart, daß $H_\alpha(u; B) \leq c\, H_{\alpha,\delta}(u; B)$ für alle Funktionen $u(x)$ gilt, für die die beiden Hölderkonstanten einen Sinn haben. Aus diesem Grunde genügt es für uns, die Konstante $H_{\alpha,\delta}(u; B)$ abzuschätzen. Nun beweisen wir den folgenden

Hilfssatz 1: Der Bereich B sei von endlich vielen, zweimal stetig differenzierbaren Hyperflächen begrenzt, die sich weder berühren noch schneiden. Die

Konstante $\lambda_\delta(B)$ sei definiert durch

$$\lambda_\delta(B) = \inf_{\substack{x^1, x^2 \in B \\ x^1 \neq x^2, |x^1 - x^2| \leq \delta}} \left[\int_{B_{x^1,x^2}} dx \, |x^1 - x^2|^{-n} \right],$$

$$B_{x^1,x^2} = \{x \ni x \in B, |x - (x^1 + x^2)/2| \leq |x^1 - x^2|/2\}.$$

δ sei dabei eine beliebig vorgegebene, positive Zahl.

Behauptung: Wenn die Funktion $u(x)$ in B meßbar ist und für alle x^0 aus B die Ungleichung

$$\int_{\substack{x \in B \\ |x - x^0| \leq \delta}} (u(x) - u(x^0))^2 \, |x - x^0|^{-n-2\alpha} \, dx \leq \mu_\delta$$

erfüllt, dann ist $u(x)$ in B gleichmäßig hölderstetig und es gilt

$$\lambda_\delta(B) \, (H_{\alpha,\delta}(u, B))^2 \leq 4 \mu_\delta.$$

Beweis: Für beliebiges x^1, x^2 und x aus B mit $|x^1 - x^2| \leq \delta$ gilt

$$(u(x^1) - u(x^2))^2 \leq 2(u(x) - u(x^1))^2 + 2(u(x) - u(x^2))^2.$$

Man integriere diese Ungleichung über den Bereich B_{x^1,x^2}:

$$(u(x^1) - u(x^2))^2 \int_{B_{x^1,x^2}} dx \leq 2 \int_{B_{x^1,x^2}} (u(x) - u(x^1))^2 \, dx + 2 \int_{B_{x^1,x^2}} (u(x) - u(x^2))^2 \, dx \leq$$

$$\leq 2 \, |x^1 - x^2|^{n+2\alpha} \times$$

$$\times \left\{ \int_{B_{x^1,x^2}} (u(x) - u(x^1))^2 |x - x^1|^{-n-2\alpha} dx + \int_{B_{x^1,x^2}} (u(x) - u(x^2))^2 |x - x^2|^{-n-2\alpha} dx \right\}.$$

Daraus folgt

$$\lambda_\delta(B) \cdot (|u(x^1) - u(x^2)|)^2 \, |x^1 - x^2|^{-2\alpha} \leq$$

$$\leq 2 \int_{x \in B, |x - x^1| \leq \delta} (u(x) - u(x^1))^2 \, |x - x^1|^{-n-2\alpha} dx +$$

$$+ 2 \int_{x \in B, |x - x^2| \leq \delta} (u(x) - u(x^2))^2 |x - x^2|^{-n-2\alpha} \, dx \leq 4 \mu_\delta$$

also

$$\lambda_\delta(B) \, (H_{\alpha,\delta}(u; B))^2 \leq 4 \mu_\delta, \qquad \text{w. z. b. w.}$$

Man sieht sehr leicht ein, daß für die von uns betrachteten Bereiche B die Konstante $\lambda_\delta(B)$ nicht verschwindet, daher folgt nunmehr die gleichmäßige Hölderstetigkeit unmittelbar.

§ 2. Integralabschätzung für die zweiten Ableitungen von u

Als nächstes beweisen wir den folgenden

Satz 3: Die symmetrische Matrix $A(x) = ((a_{ik}(x)))$ habe in der Kugel $|x|^2 = \sum_{i=1}^{n} x_i^2 \leq r_0^2$ meßbare Koeffizienten und genüge für jedes x aus $|x| \leq r_0$ einer K_ε-Bedingung mit einem von x unabhängigen ε. Ferner gelte $\sum_{i=1}^{n} a_{ii}(x) > p$ mit einer positiven Konstanten p, die ebenfalls nicht von x abhängen möge. Die Funktion $u(x)$ sei in $|x| \leq r_0$ stetig differenzierbar und habe stückweise

stetige zweite Ableitungen. m sei eine beliebige Zahl $\geqq 4$. Dann gibt es Konstanten c_1, c_2, die nur von ε, p, n, m abhängen, derart, daß die Abschätzung

$$\int\limits_{r\leqq r_0} \sum_{i,k=1}^{n} (\partial^2 u/\partial x_i\,\partial x_k)^2 r^{3-n}\,|r_0^2-r^2|^m dx \leqq c_1 \int\limits_{r\leqq r_0} \left(\sum_{i,k=1}^{n} a_{ik}\frac{\partial^2 u}{\partial x_i\,\partial x_k}\right)^2 \times$$
$$\times\, r^{3-n}(r_0^2-r^2)^m dx + c_2 r_0^{2m-2}\int\limits_{r\leqq r_0} u^2 r^{1-n} dx$$

gilt.

Der Beweis dieses Satzes gründet sich im wesentlichen auf folgenden

Hilfssatz 2: Wenn $u(x)$ in $|x|\leqq r_0$ stetig differenzierbar ist und stückweise stetige zweite Ableitungen hat, dann gilt für jedes $\varkappa>0$ und jedes $m\geqq 4$ sowie für $n\geqq 3$ eine Abschätzung der Form

$$(1-\varkappa)\int\limits_{r\leqq r_0} \sum_{i,k=1}^{n} (\partial^2 u/\partial x_i\,\partial x_k)^2 r^{3-n}(r_0^2-r^2)^m dx - \int\limits_{r\leqq r_0} (\Delta u)^2 r^{3-n}(r_0^2-r^2)^m dx \leqq$$
$$\leqq K_1(\varkappa)\, r_0^{2m-2}\int\limits_{r\leqq r_0} u^2 r^{1-n} dx\,,$$

wo $K_1(\varkappa)$ nur von $\varkappa$, n und m, nicht aber von u und r_0 abhängt.

Beweis: Wir gehen von der Identität

$$(\Delta u)^2 - \sum_{i,k=1}^{n} (\partial^2 u/\partial x_i\,\partial x_k)^2 = \sum_{i,k=1}^{n} \partial^2/\partial x_i\,\partial x_k(\partial u/\partial x_i\,\partial u/\partial x_k) - \Delta\left(\sum_{i=1}^{n}(\partial u/\partial x_i)^2\right)$$

aus. Durch Multiplikation mit $(r_0^2-r^2)^m r^{3-n}$ und Integration über die Kugel $r\leqq r_0$ sowie partielle Integration folgt[3])

$$J = \int\limits_{r\leqq r_0}\left(\sum_{i,k=1}^{n}\partial^2 u/\partial x_i\,\partial x_k\right)^2 r^{3-n}(r_0^2-r^2)^m dx - \int\limits_{r\leqq r_0}(\Delta u)^2 r^{3-n}(r_0^2-r^2)^m dx$$
$$= \int\limits_{r\leqq r_0}\Bigg[\sum_{i=1}^{n}(\partial u/\partial x_i)^2 \Delta(r^{3-n}(r_0^2-r^2)^m) -$$
$$- \sum_{i,k=1}^{n}\partial u/\partial x_i\,\partial u/\partial x_k(\partial^2/\partial x_i\,\partial x_k(r^{3-n}(r_0^2-r^2)^m))\Bigg]dx\,.$$

Wegen des Faktors $(r_0^2-r^2)^m$ treten dabei keine Randglieder auf. Nun gilt

$$\Delta(r^{3+n}(r_0^2-r^2)^m) = (3-n)\,r^{1-n}(r_0^2-r^2)^m +$$
$$+\,2m(n-6)\,r^{3-n}(r_0^2-r^2)^{m-1} + 4m(m-1)\,r^{5-n}(r_0^2-r^2)^{m-2}$$

und

$$\partial^2/\partial x_i\,\partial x_k(r^{3-n}(r_0^2-r^2)^m)$$
$$= \{(3-n)\,r^{1-n}\delta_{ik} + (3-n)(1-n)\,r^{-1-n}x_i x_k\}(r_0^2-r^2)^m -$$
$$-\,2m\,r^{1-n}\{2(3-n)\,x_i x_k + r^2\delta_{ik}\}(r_0^2-r^2)^{m-1} +$$
$$+\,4m(m-1)\,r^{3-n}x_i x_k(r_0^2-r^2)^{m-2};$$

---

[3]) Wie in [1] kann man leicht einsehen, daß das Auftreten von Sprüngen in den zweiten Ableitungen von $u(x)$ für die partielle Integration keine Bedeutung hat, weil die entsprechenden Terme sich wegheben.

somit wird der Integrand der rechten Seite gleich

$$
\begin{aligned}
&-(n-3)(n-1)\,r^{1-n}(r_0^2-r^2)^m(\partial u/\partial r)^2+\\
&+2m(n-5)\,r^{3-n}(r_0^2-r^2)^{m-1}\sum_{i=1}(\partial u/\partial x_i)^2-\\
&-4m(n-3)\,r^{3-n}(r_0^2-r^2)^{m-1}(\partial u/\partial r)^2+\\
&+4m(m-1)\,r^{5-n}(r_0^2-r^2)^{m-2}\left(\sum_{i=1}^{m}(\partial u/\partial x_i)^2-(\partial u/\partial r)^2\right).
\end{aligned}
$$

Dies kann für $n>2$ nach oben durch

$$
[2m(n-5)(r_0^2-r^2)+4m(m-1)r^2]\,r^{3-n}(r_0^2-r^2)^{m-2}\left(\sum_{i=1}^{n}(\partial u/\partial x_i)^2-(\partial u/\partial r)^2\right)
$$

abgeschätzt werden. Somit folgt

$$
\begin{aligned}
J &\leqq c r_0^2\int\limits_{r\leqq r_0}\left(\sum_{i=1}^{n}(\partial u/\partial x_i)^2-(\partial u/\partial r)^2\right)r^{3-n}(r_0^2-r^2)^{m-2}\,dx\\
&=-c r_0^2\int\limits_{r\leqq r_0}u(\Delta u-\partial^2 u/\partial r^2-(n-1)/r\;\partial u/\partial r)\,r^{3-n}(r_0^2-r^2)^{m-2}\,dx
\end{aligned}
$$

mit

$$
c=r_0^{-2}\operatorname*{Max}_{r\leqq r_0}[2m(n-5)(r_0^2-r^2)+4m(m-1)r^2]\,.
$$

Offenbar hängt c nicht von r_0 ab, sondern ausschließlich von n und m.

Es folgt endlich

$$
\begin{aligned}
\left|\int\limits_{r\leqq r_0}u\,\partial u/\partial r\;r^{2-n}(r_0^2-r^2)^{m-2}\,dx\right| &= \left|-1/2\int\limits_{r\leqq r_0}u^2r^{1-n}\,\partial/\partial r\,[r(r_0^2-r^2)^{m-2}]\,dx\right|\leqq\\
&\leqq c' r_0^{2m-4}\int\limits_{r\leqq r_0}u^2r^{1-n}\,dx\,,
\end{aligned}
$$

da wegen $m\geqq 4$ wiederum der Randterm verschwindet. Die Konstante

$$
c'=\operatorname*{Max}_{r\leqq r_0}\{r_0^{4-2m}\,\partial/\partial r\,[r(r_0^2-r^2)^{m-2}]\}
$$

hängt offenbar nur von m und n, insbesondere aber nicht von r_0 ab. Daraus folgt zusammen

$$
\begin{aligned}
J\leqq&-c r_0^2\int\limits_{r\leqq r_0}u\,\Delta u\,r^{3-n}(r_0^2-r^2)^{m-2}dx+c r_0^2\int\limits_{r\leqq r_0}u\,\partial^2u/\partial r^2\,r^{3-n}(r_0^2-r^2)^{m-2}dx+\\
&+cc' r_0^{2m-2}\int\limits_{r\leqq r_0}u^2r^{1-n}dx\leqq\\
\leqq&\;\varkappa_1 c r_0^2\int\limits_{r\leqq r_0}\left[(\Delta u)^2+\sum_{i,k=1}^{n}(\partial^2u/\partial x_i\,\partial x_k)^2\right]r^{3-n}(r_0^2-r^2)^{2m-4}dx+\\
&+(2\varkappa_1)^{-1}c r_0^4\int\limits_{r\leqq r_0}u^2r^{1-n}dx+cc' r_0^{2m-2}\int\limits_{r\leqq r_0}u^2r^{1-n}dx\,.
\end{aligned}
$$

Wegen $m \geqq 4$ folgt $2m - 4 \geqq m$, daher gilt abschließend

$$J \leqq \varkappa_1 c r_0^{2m-6} \int\limits_{r \leqq r_0} \left[(\Delta u)^2 + \sum_{i,k=1}^{n} (\partial^2 u/\partial x_i\, \partial x_k)^2 \right] r^{3-n} (r_0^2 - r^2)^m \, dx +$$

$$+ \left((2\varkappa_1)^{-1} c r_0^4 + c c' r_0^{2m-2} \right) \int\limits_{r \leqq r_0} u^2 r^{1-n} dx \, .$$

Man setze noch $\varkappa_2 = \varkappa_1 c r_0^{2m-6}$, dann wird der Term vor dem letzten Integral zu

$$(2\varkappa_2)^{-1} c^2 r_0^{2m-2} + c c' r_0^{2m-2} \, .$$

Nunmehr folgt unmittelbar die Behauptung mit

$$\varkappa = 2\varkappa_2 (1 + \varkappa_2)^{-1} \, , \quad K_2 = c(1 + \varkappa_2)^{-1} ((2\varkappa_2)^{-1} c + c') \, .$$

Der Fall $n = 2$ erfordert eine Sonderbehandlung. Man kann dann ganz ähnliche Betrachtungen anstellen, die zu der wesentlichen Aussage des Satzes 7 führen. Wir wollen dies jedoch hier übergehen.

Nun wollen wir Satz 3 beweisen. Man setze $\sum_{i,k=1}^{n} a_{ik} \partial^2 u/\partial x_i\, \partial x_k = f$. Aus dem soeben bewiesenen Hilfssatz folgt sofort für beliebiges $\eta > 0$ die Abschätzung

$$\eta K_1(\varkappa)\, r_0^{2m-2} \int\limits_{r \leqq r_0} u^2 r^{1-n} dx + \int\limits_{r \leqq r_0} f^2 (r_0^2 - r^2)^m r^{3-n} dx \geqq$$

$$\geqq \int\limits_{r \leqq r_0} \sum_{i,j,k,l=1}^{n} (a_{ij} a_{kl} + \eta(1 - \varkappa)\, \delta_{ik} \delta_{jl} - \eta \delta_{ij} \delta_{kl})\, \partial^2 u/\partial x_i\, \partial x_j \times$$

$$\times\, \partial^2 u/\partial x_k\, \partial x_l (r_0^2 - r^2)^m r^{3-n} dx \, .$$

Unser Bestreben geht dahin zu zeigen, daß für hinreichend kleine Wahl von $\varkappa$ und η die $n^2 \times n^2$-Matrix

$$((b_{ij,kl})) = ((a_{ij} a_{kl} + \eta(1 - \varkappa)\, \delta_{ik} \delta_{jl} - \eta \delta_{ij} \delta_{kl})) \, ,$$

wo i und j die Zeilenindizes, k und l die Spaltenindizes sein mögen, positiv definit ist, und zwar derart, daß eine positive untere Schranke p_0 angegeben werden kann, die allein vom ε der K_ε-Bedingung sowie von p, n, $\varkappa$ und η abhängt. Dies führt ersichtlich sofort auf die in Satz 3 behauptete Abschätzung mit $c_1 = p_0^{-1}$, $c_2 = \eta K_1(\varkappa) p_0^{-1}$. Die Konstanten c_1, c_2 können so gewählt werden, daß sie allein von ε, p, n und m abhängen, falls man zeigt, daß die Auswahl von p_0, $\varkappa$ und η bereits nach Vorgabe von ε, p, n und m vorgenommen werden kann.

Wir führen im n^2-dimensionalen Raum aller reellen $n \times n$-Matrizen ein Skalarprodukt ein durch

$$(F, G) = \sum_{i,k=1}^{n} f_{ik} g_{ik} = \mathrm{Spur}\,(F G) \, , \quad F = ((f_{ik})) \, , \quad G = ((g_{ik})) \, .$$

Die Ungleichung

$$\sum_{i,j,k,l=1}^{n} b_{ij,kl} z_{ij} z_{kl} \geqq p_0 \sum_{i,j,k,l=1}^{n} z_{ij}^2$$

kann dann auch wie folgt geschrieben werden:

$$[(A, Z)]^2 - \eta\,[(E, Z)]^2 + ((1-\varkappa)\,\eta - p_0)\,(Z, Z) \geq 0 \quad \text{für alle} \quad Z = ((z_{ik})) .$$

Dabei sei $E = ((\delta_{ik}))$ gesetzt. Offenbar gilt $E = n^{-1/2} E_0$, wo $(E_0, E_0) = 1$ ist; ferner kann A zerlegt werden in der Form

$$A = \sigma E_0 + \tau A_0$$

mit

$$(A_0, A_0) = 1\,, \quad (A_0, E_0) = 0\,.$$

Die Produkte (E_0, Z), (A_0, Z), als die Komponenten von Z in Richtung E_0 bzw. A_0, seien im ferneren mit z_{E_0} bzw. z_{A_0} bezeichnet. Dann hat man

$$(Z, Z) \geq z_{E_0}^2 + z_{A_0}^2\,,$$

und es folgt

$$\begin{aligned}(A, Z)^2 - \eta\,(E, Z)^2 + (\eta\,(1-\varkappa) - p_0)\,(Z, Z) &\geq \\ \geq (\sigma z_{E_0} + \tau z_{A_0})^2 - n\,\eta\, z_{E_0}^2 &+ (1-\varkappa_1)\,\eta\,(z_{E_0}^2 + z_{A_0}^2) \\ = (\sigma^2 - n\,\eta + (1-\varkappa_1)\,\eta)\, z_{E_0}^2 &+ (\tau^2 + (1-\varkappa_1)\,\eta)\, z_{A_0}^2 + 2\sigma\tau z_{E_0} z_{A_0}\,.\end{aligned}$$

Dabei haben wir $p_0 = \eta\varkappa'$ gesetzt und $\varkappa_1 = \varkappa + \varkappa'$ eingeführt, so daß der Ausdruck $(1-\varkappa)\,\eta - p_0$ in $(1-\varkappa_1)\,\eta$ übergeht. So erhalten wir also eine quadratische Form in z_{E_0} und z_{A_0}; als notwendige und hinreichende Bedingung dafür, daß diese ≥ 0 ist, ergibt sich daher

$$\sigma^2 - n\,\eta + (1-\varkappa_1)\,\eta \geq 0\,, \qquad \tau^2 + (1-\varkappa_1) \geq 0$$
$$(\sigma^2 - n\,\eta + (1-\varkappa_1)\,\eta)\,(\tau^2 + (1-\varkappa_1)\,\eta) - \sigma^2\tau^2 \geq 0\,.$$

Offenbar ist die zweite Bedingung für $\varkappa_1 < 1$, $\eta > 0$ von selbst erfüllt, die dritte Bedingung schreiben wir folgendermaßen um:

$$(1-\varkappa_1 - n)\,(1-\varkappa_1)\,\eta + (1-\varkappa_1)\sigma^2 - (n - 1 + \varkappa_1)\tau^2 \geq 0\,.$$

Für kleines η und kleines $\varkappa_1$ bedeutet dies

$$(n-1)\tau^2 \leq (1-\varepsilon_1)\sigma^2 - \eta_1 \;\text{ mit }\; \varepsilon_1 = 1 - (n-1)\,(1-\varkappa_1)/(n-1+\varkappa_1),$$
$$\eta_1 = \eta\,(1-\varkappa_1)\,(n-1)\,.$$

Endlich folgt

$$\sigma = (E_0, A) = n^{-1/2}\,\text{Spur}\,(A) = n^{-1/2} \sum_{i=1}^{n} \lambda_i$$

$$\begin{aligned}\tau^2 &= (A - 1/n\,(\text{Spur}\,(A))\,E,\; A - 1/n\,(\text{Spur}\,(A))\,E) \\ &= \text{Spur}\,\{A - 1/n\,(\text{Spur}\,(A))\,E\}^2 = \sum_{i=1}^{n} (\lambda_i - s)^2\,,\end{aligned}$$

wobei λ_i die Eigenwerte von A bezeichnen mögen und wo s das arithmetische Mittel der Eigenwerte sei: $s = 1/n \sum_{i=1}^{n} \lambda_i$.

Daher ist obige Abschätzung gleichbedeutend mit

$$(n-1) \sum_{i=1}^{n} (\lambda_i - s)^2 \leq 1/n\,(1-\varepsilon_1)\left(\sum_{i=1}^{n} \lambda_i\right)^2 - \eta_1\,.$$

Nun wurde in den Voraussetzungen von Satz 3 für die Matrix $((a_{ik}))$ eine K_ε-Bedingung gefordert. Daher folgt

$$(n-1)\sum_{i=1}^{n}(\lambda_i-s)^2=(n-1)\sum_{i=1}^{n}\lambda_i^2-(n-1)/n\left(\sum_{i=1}^{n}\lambda_i\right)^2$$
$$=(n-1)\sum_{i,k=1}^{n}a_{ik}^2-(n-1)/n\left(\sum_{i=1}^{n}a_{ii}\right)^2\leqq[(n-\varepsilon)/n-(n-1)/n]\left(\sum_{i=1}^{n}a_{ii}\right)^2\leqq$$
$$\leqq 1/n\,(1-\varepsilon)\left(\sum_{i=1}^{n}\lambda_i\right)^2.$$

Überdies war vorausgesetzt, daß

$$0<p\leqq\sum_{i=1}^{n}\lambda_i=\sum_{i=1}^{n}a_{ii}$$

gleichmäßig in x gelte. Aus beidem zusammen folgt sofort

$$(n-1)\sum_{i=1}^{n}(\lambda_i-s)^2\leqq 1/n\,(1-\varepsilon/2)\left(\sum_{i=1}^{n}\lambda_i\right)^2-p^2\varepsilon/(2n).$$

Für

$$\varepsilon_1\leqq\varepsilon/2\,,\quad \eta_1\leqq p^2\varepsilon/(2n)\,,$$

d. h., für

$$1-(n-1)\,(1-\varkappa_1)/(n-1+\varkappa_1)\leqq\varepsilon/2\,,\quad (n-1)\,\eta\,(1-\varkappa_1)\leqq p^2\varepsilon/(2n)$$

ist daher die dritte Bedingung erfüllt.

Endlich läßt sich die erste Bedingung offenbar durch die Forderung

$$(n-1+\varkappa_1)\,\eta\leqq p^2$$

erfüllen, denn sie ist gleichbedeutend mit

$$(n-1+\varkappa_1)\,\eta\leqq\left(\sum_{i=1}^{n}a_{ii}\right)^2.$$

Somit, wenn $\varkappa_1$ aus

$$1-(n-1)\,(1-\varkappa_1)/(n-1+\varkappa_1)=\varepsilon/2$$

bestimmt wird und alsdann

$$\eta=\mathrm{Min}\,[p^2\varepsilon/[(n-1)\,(1-\varkappa)],\,p^2/(n-1+\varkappa_1)]$$

gesetzt wird, so ergeben sich Konstanten, die alle verlangten Bedingungen erfüllen.

Endlich möge $\varkappa=\varkappa'=\varkappa_1/2$ und $p_0=\eta\varkappa'$ gesetzt werden, dann ergibt sich die in Satz 3 verlangte Ungleichung, und der Satz ist bewiesen.

§ 3. Integralabschätzung der ersten Ableitungen

Als Verbindungsglied zwischen Hilfssatz 1 und Satz 3 benötigen wir noch zwei weitere kleine Hilfssätze.

Hilfssatz 3: Die Funktion $y(t)$ der einen unabhängigen Veränderlichen t sei in $0 \leq t \leq 1$ zweimal stetig differenzierbar. Dann gilt

$$\int_0^1 (dy/dt)^2(1-t)^2 dt \leq \pi^2(1 + 5/2\,\pi^2) \int_0^1 (d^2y/dt^2)^2 t^2(1-t)^4 dt + $$
$$+ 5(1/2 + \pi^2) \int_0^1 y^2 dt \,.$$

Beweis: Es folgt

$$0 \leq \int_0^1 (d^2y/dt^2 \sin 2\pi t + \alpha\, dy/dt)^2 \cos^2 \pi/2 t\, dt$$
$$= \int_0^1 (d^2y/dt^2)^2 \sin^2 2\pi t \cos^2 \pi/2\, t\, dt + \alpha^2 \int_0^1 (dy/dt)^2 \cos^2 \pi/2\, t\, dt +$$
$$+ \alpha \int_0^1 d/dt((dy/dt)^2) \cos^2 \pi/2\, t \sin 2\pi t\, dt \,.$$

Das letzte Integral kann durch partielle Integration umgeformt werden in

$$-\alpha \int_0^1 (dy/dt)^2(-\pi/2 \sin \pi t \sin 2\pi t + 2\pi \cos^2 \pi/2\, t \cos 2\pi t)\, dt \,.$$

Daher ergibt sich wegen $\cos 2\pi t = 1 - 2\sin^2 \pi t$ folgende Ungleichung

$$(2\pi\alpha - \alpha^2) \int_0^1 (dy/dt)^2 \cos^2 \pi/2\, t\, dt \leq \int_0^1 (d^2y/dt^2)^2 \sin^2 2\pi t \cos^2 \pi/2\, t\, dt +$$
$$+ 4\pi\alpha \int_0^1 (dy/dt)^2 \cos^2 \pi/2\, t \sin^2 \pi t\, dt + \alpha\pi/2 \int_0^1 (dy/dt)^2 \sin \pi t \sin 2\pi t\, dt \,.$$

Wegen $|\sin 2\pi t| \leq 2|\sin \pi t|$, $\cos^2 \pi/2\, t \leq 1$, sind die letzten beiden Integrale kleiner oder gleich

$$5\pi\alpha \int_0^1 (dy/dt)^2 \sin^2 \pi t\, dt \,,$$

und dieses Integral kann wie folgt weiter abgeschätzt werden:

$$5\pi\alpha \int_0^1 (dy/dt)^2 \sin^2 \pi t\, dt = -5\pi\alpha \int_0^1 y\, d^2y/dt^2 \sin^2 \pi t\, dt -$$
$$- 5\pi^2\alpha \int_0^1 y\, dy/dt \sin 2\pi t\, dt \leq$$
$$\leq 5/2\pi\alpha \int_0^1 (d^2y/dt^2)^2 \sin^4 \pi t\, dt + 5/2\,\pi\alpha \int_0^1 y^2 dt +$$
$$+ 5\pi^3\alpha \int_0^1 y^2 \cos 2\pi t\, dt \,.$$

Zusammenfassend hat man so für $\alpha = \pi$ wegen $\sin^2\pi t \leqq 4\cos^2\pi/2t$ und $\cos 2\pi t \leqq 1$ die Abschätzung

$$\pi^2 \int_0^1 (dy/dt)^2 \cos^2 \pi/2\, t\, dt \leqq (4 + 10\pi^2) \int_0^1 (d^2y/dt^2) \sin^2\pi t \cos^2 \pi/2\, t\, dt +$$

$$+ (5/2\pi^2 + 5\pi^4) \int_0^1 y^2 dt\,.$$

Hieraus folgt die in Hilfssatz 3 ausgesprochene Ungleichung unmittelbar wegen

$$(1-t)^2 \leqq \cos^2 \pi/2\, t \leqq \pi^2/4\, (1-t)^2\,, \quad \sin^2\pi t \leqq \pi^2 t^2 (1-t)^2\,.$$

Hilfssatz 4: Sei $y(t)$ zweimal stetig differenzierbar in $0 \leqq t \leqq 1$ und $y(0) = 0$.
Behauptung:

$$\int_0^1 y^2\, dt/t^2 \leqq 4 \int_0^1 (dy/dt)^2 (1-t)^2 dt\,.$$

Beweis: Es ist

$$0 \leqq \int_0^1 ((1-t)\, dy/dt - (2t)^{-1} y)^2 dt$$

$$= \int_0^1 (dy/dt)^2 (1-t)^2 dt + 1/4 \int_0^1 y^2\, dt/t^2 - 1/2 \int_0^1 [d/dt(y^2)]\, (t^{-1} - 1)\, dt\,.$$

Partielle Integration liefert

$$-1/2 \int_0^1 [d/dt(y^2)]\, (t^{-1}-1)\, dt = -1/2 \int_0^1 y^2\, dt/t^2\,,$$

da die auftretenden Randterme verschwinden. Daraus folgt unmittelbar die behauptete Ungleichung.

Nun ergibt sich sofort

Satz 4: Die Funktion $u(x)$ sei stetig differenzierbar und habe stückweise stetige zweite Ableitungen in $|x| \leqq r_0$.

Behauptung: Es gibt Konstanten c_4 und c_5, die nicht von u und nicht von r_0, sondern nur von n abhängen, derart, daß gilt

$$\int_{r \leqq r_0} \sum_{i=1}^{n} (\partial u/\partial x_i)^2 r^{1-n} (r_0^2 - r^2)^4 dx \leqq$$

$$\leqq c_4 \int_{r \leqq r_0} \sum_{i,k=1}^{n} (\partial^2 u/\partial x_i \partial x_k)^2\, r^{3-n} (r_0^2 - r^2)^4 dx + c_5 r_0^6 \int_{r \leqq r_0} u^2 r^{1-n} dx\,.$$

Beweis: Um die angegebene Abschätzung zu gewinnen, benutzen wir die folgende Ungleichung:

$$\int_{r \leqq r_1} v^2 do_1 \leqq (n-1)^{-1} r_1^2 \int_{r = r_1} \sum_{k=1}^{n} (\partial v/\partial x_k)^2 do_1 + (\omega_n)^{-1} \Bigl(\int_{r = r_1} v\, do_1\Bigr)^2 .$$

Darin ist do_1 das Flächenelement, ω_n die Oberfläche der n-dimensionalen Einheitskugel, $v(x)$ sei eine beliebige auf der Kugel $|x| = r_1$ zweimal stetig differenzierbare Funktion. Diese Abschätzung kommt folgendermaßen zustande: Es ist bekanntlich

$$\int_{r=r_1} \left(\sum_{k=1}^{n} (\partial v/\partial x_k)^2 - (\partial v/\partial r)^2 \right) do_1 = -r_1^{-2} \int_{r=r_1} v\, \Delta^* v\, do_1 ,$$

wo $\Delta^* = r^2 \Delta - (n-1)\, r\, \partial/\partial r - r^2\, \partial^2/\partial r^2$ der Beltramische Operator auf der Einheitskugel sei. Bekanntlich besitzt nun der Operator $-\Delta^*$ die Kugelfunktionen Y_l^n, $l = 0, 1, 2, \ldots$, als Eigenfunktionen zum Eigenwert $l(l+n-2)$, und es läßt sich daraus ein vollständiges, orthogonales und normiertes System von Eigenfunktionen bilden. Der Eigenwert Null zu $l = 0$ besitzt die Vielfachheit eins, und alle zugehörigen Eigenfunktionen sind Vielfache der Funktion $Y_0^n = 1$. Da alle weiteren Eigenwerte positiv und sogar größer oder gleich $n-1$ sind, folgt direkt

$$\int_{r=r_1} v^2\, do_1 \leqq -(n-1)^{-1} \int_{r=r_1} v\Delta^* v\, do_1 + (\omega_n)^{-1} \left(\int_{r=r_1} v\, do_1 \right)^2 .$$

Man setze in diese Ungleichung nun der Reihe nach $v = \partial u/\partial x$, $i = 1, \ldots, n$ und summiere über i:

$$\int_{r=r_1} \sum_{i=1}^{n} (\partial u/\partial x_i)^2\, do_1 \leqq$$

$$\leqq (n-1)^{-1} r_1^2 \int_{r=r_1} \sum_{i,k=1}^{n} (\partial^2 u/\partial x_i\, \partial x_n)^2\, do_1 + (\omega_n)^{-1} \sum_{i=1}^{n} \left(\int_{r=r_1} \partial u/\partial x_i\, do_1 \right)^2 .$$

Die Integrale $\int_{r=r_1} \partial u/\partial x_i\, do_1$ berechnen wir folgendermaßen: Es ist

$$\int_{r \leqq r_1} \partial u/\partial x_i\, dx = r_1^{n-1} \int_{r=r_1} u\, x_i/r\, do_1 ,$$

also folgt durch Differentiation:

$$r_1^{n-1} \int_{r=r_1} \partial u/\partial x_i\, do_1 = (n-1)\, r_2^{n-1} \int_{r=r_1} u\, x_i/r^2\, do_1 + r_1^{n-1} \int_{r=r_1} \partial u/\partial r\, x_i/r\, do_1 .$$

Also folgt

$$\int_{r=r_1} \partial u/\partial x_i\, do_1 = \int_{r=r_1} (\partial u/\partial r + (n-1)/r\, u)\, x_i/r\, do_1 .$$

Mit Hilfe dieser Beziehung schätzen wir oben weiter ab.

$$\int_{r=r_1} \sum_{i=1}^{n} (\partial u/\partial x_i)^2\, do_1 \leqq (n-1)^{-1} r_1^2 \int_{r=r_1} \sum_{i,k=1}^{n} (\partial^2 u/\partial x_i\, \partial x_k)^2\, do_1 +$$

$$+ \int_{r=r_1} (\partial u/\partial r + (n-1)/r\, u)^2\, do_1 \leqq (n-1)^{-1} r_1^2 \int_{r=r_1} \sum_{i,k=1}^{n} (\partial^2 u/\partial x_i\, \partial x_k)^2\, do_1 +$$

$$+ 2 \int_{r=r_1} (\partial u/\partial r)^2\, do_1 + 2(n-1)^2 \int_{r=r_1} u^2/r^2\, do_1 .$$

Diese Ungleichung multiplizieren wir mit $(r_0^2 - r_1^2)^4$ und integrieren nach dr_1:

$$\int\limits_{r \leq r_0} \sum_{i=1}^{n} (\partial u/\partial x_i)^2 r^{1-n} (r_0^2 - r^2)^4 dx \leq (n-1)^{-1} \int\limits_{r \leq r_0} \sum_{i,k=1}^{n} (\partial^2 u/\partial x_i\, \partial x_k)^2 \times$$

$$\times r^{3-n} (r_0^2 - r^2)^4\, dx +$$

$$+ 2 r_0^4 \int\limits_{r = r_0} (\partial u/\partial r)^2 r^{1-n} (r_0^2 - r^2)^2 dx + 2(n-1)^2 r_0^8 \int\limits_{r \leq r_0} u^2 r^{-1-n} dx \leq$$

$$\leq (n-1)^{-1} \int\limits_{r \leq r_0} \sum_{i,k=1}^{n} (\partial^2 u/\partial x_i\, \partial x_k)^2 r^{3-n} (r_0^2 - r^2)^4 dx +$$

$$+ (2 + 8(n-1)^2)\, r_0^4 \int\limits_{r \leq r_0} (\partial u/\partial r)^2 r^{1-n} (r_0^2 - r^2)^2 dx .$$

Zur Durchführung des letzten Schrittes wurde dabei Hilfssatz 4 verwendet. Nämlich, aus Hilfssatz 4 folgt für die Funktion $y(t) = u(tx)$ bei festem x mit $|x| = r_0$ die Abschätzung

$$\int_0^{r_0} u^2\, dr/r^2 \leq 4 r_0^{-4} \int_0^{r_0} (\partial u/\partial r)^2\, (r_0^2 - r^2)^2\, dr,$$

welche nach Integration über die Einheitskugel die Ungleichung

$$\int\limits_{r \leq r_0} u^2 r^{-1-n}\, dx \leq 4 r_0^{-4} \int\limits_{r \leq r_0} (\partial u/\partial r)^2 r^{1-n} (r_0^2 - r^2)^2\, dx$$

ergibt, welche zur Abschätzung im obigen Sinne benutzt werden kann.

Endlich werde Hilfssatz 3 auf die Funktion $y(t) = u(tx)$, $0 \leq t \leq 1$ (bei festem x mit $|x| = r_0$) angewandt. Dann folgt

$$\int_0^{r_0} (\partial u/\partial r)^2 (r_0 - r)^2\, dr \leq \pi^2 (1 + 5/2\pi^2)\, r_0^{-2} \int_0^{r_0} (\partial^2 u/\partial r^2)^2 r^2 (r_0 - r)^4\, dr +$$

$$+ 5(\pi^2 + 1/2) \int_0^{r_0} u^2\, dr .$$

Dies werde über die Kugel $|x| = r_0$ integriert:

$$\int\limits_{r \leq r_0} (\partial u/\partial r)^2 r^{1-n} (r_0 - r)^2 dx \leq \pi^2 (1 + 5/2\, \pi^2)\, r_0^{-2} \int\limits_{r \leq r_0} (\partial^2 u/\partial r^2)^2 r^{3-n} (r_0 - r)^4\, dx +$$

$$+ 5(\pi^2 + 1/2) \int\limits_{r \leq r_0} u^2 r^{1-n}\, dx .$$

Wir führen darin $r_0^2 - r^2$ ein anstatt $r_0 - r$ und erhalten

$$\int\limits_{r \leq r_0} (\partial u/\partial r)^2 r^{1-n} (r_0^2 - r^2)^2 dx \leq 4\pi^2 (1 + 5/2\, \pi^2)\, r_0^{-4} \int\limits_{r \leq r_0} (\partial^2 u/\partial r^2)^2 r^{3-n} (r_0^2 - r^2)^4\, dx$$

$$+ 20(\pi^2 + 1/2)\, r_0^2 \int\limits_{r \leq r_0} u^2 r^{1-n}\, dx .$$

Wegen $\partial^2 u/\partial r^2 = \sum_{i,k=1}^{n} x_i/r\, x_k/r\, \partial^2 u/\partial x_i\, \partial x_k$ folgt $(\partial^2 u/\partial r^2)^2 \leq \sum_{i,k=1}^{n} (\partial^2 u/\partial x_i\, \partial x_k)^2$, also kann man weiter abschätzen:

$$\leq 4\pi^2(1 + 5/2\,\pi^2)\, r_0^{-4} \int_{r \leq r_0} \sum_{i,k=1}^{n} (\partial^2 u/\partial x_i\, \partial x_k)^2 r^{3-n} (r_0^2 - r^2)^4\, dx +$$
$$+ 20(\pi^2 + 1/2)\, r_0^2 \int_{r \leq r_0} u^2 r^{1-n}\, dx\,.$$

Aus beiden Abschätzungen zusammen folgt endlich

$$\int_{r \leq r_0} \sum_{i=1}^{n} (\partial u/\partial x_i)^2 r^{1-n} (r_0^2 - r^2)^4\, dx \leq$$
$$\leq (n-1)^{-1} \int_{r \leq r_0} \sum_{i,k=1}^{n} (\partial^2 u/\partial x_i\, \partial x_k)^2 r^{3-n} (r_0^2 - r^2)^4\, dx +$$
$$+ 2(1 + 4(n-1)^2)\, r_0^4 \int_{r \leq r_0} (\partial u/\partial r)^2 r^{1-n} (r_0^2 - r^2)^2\, dx \leq$$
$$\leq c_4 \int_{r \leq r_0} \sum_{i,k=1}^{n} (\partial^2 u/\partial x_i\, \partial x_k)^2 r^{3-n} (r_0^2 - r^2)^4\, dx + c_5 r_0^6 \int_{r \leq r_0} u^2 r^{1-n}\, dx$$

mit

$$c_4 = (n-1)^{-1} + 8\pi^2(1 + 5/2\,\pi^2)(1 + 4(n-1)^2)$$
$$c_5 = 40(\pi^2 + 1/2)(1 + 4(n-1)^2)\,,$$

also die Aussage von Satz 4.

Kombination von Satz 3 und Satz 4 liefert nun sofort

Satz 5: Unter den Voraussetzungen von Satz 3 gibt es Konstanten c_6 und c_7, die nur von ε, p und n, nicht aber von r_0, u oder den $a_{ik}(x)$ abhängen, derart, daß gilt

$$\int_{r \leq r_0} \sum_{i=1}^{n} (\partial u/\partial x_i)^2 r^{1-n} (r_0^2 - r^2)^4\, dx \leq c_6 \int_{r \leq r_0} \Big(\sum_{i,k=1}^{n} a_{ik}\, \partial^2 u/\partial x_i\, \partial x_k \Big)^2 \times$$
$$\times r^{3-n} (r_0^2 - r^2)^4\, dx +$$
$$+ r_0^6 \int_{r \leq r_0} u^2 r^{1-n}\, dx\,.$$

Hieraus folgt weiterhin

Satz 6: Sei

$$L(u) = \sum_{i,k=1}^{n} a_{ik}(x)\, \partial^2 u/\partial x_i\, \partial x_k + \sum_{i=1}^{n} b_i(x)\, \partial u/\partial x_i + b_0(x) u$$

ein linearer Differentialausdruck, für dessen Hauptteil $\sum_{i,k=1}^{n} a_{ik}(x)\, \partial^2 u/\partial x_i\, \partial x_k$ die Voraussetzungen von Satz 3 erfüllt sein mögen. Es gebe ferner eine Konstante M, derart, daß gilt

$$\sum_{i=0}^{n} (b_i(x))^2 \leq M\,.$$

Dann gibt es ein $r_1 > 0$ und Konstanten c_9, c_{10}, welche nur von ε, p, n und M abhängen, so daß für $r_0 \leq r_1$ gilt

$$\int\limits_{r \leq r_0} \sum_{i=1}^{n} (\partial u/\partial x_i)^2 r^{1-n} (r_0^2 - r^2)^4 dx \leq c_9 \int\limits_{r \leq r_0} (L(u))^2 r^{3-n} (r_0^2 - r^2)^4 dx + \\ + c_{10} \int\limits_{r \leq r_0} u^2 r^{1-n} dx .$$

Beweis: Nach Satz 5 folgt für beliebiges r_0 die Ungleichung

$$\int\limits_{r \leq r_0} \sum_{i=1}^{n} (\partial u/\partial x_i)^2 r^{1-n} (r_0^2 - r^2)^4 dx \leq 2c_6 \int\limits_{r \leq r_0} (L(u))^2 r^{3-n} (r_0^2 - r^2)^4 dx + \\ + 2c_6 M r_0^2 \int\limits_{r \leq r_0} \sum_{i=1}^{n} (\partial u/\partial x_i)^2 r^{1-n} (r_0^2 - r^2)^4 dx + \\ + 2c_6 M r_0^{10} \int\limits_{r \leq r_0} u^2 r^{1-n} dx + c_7 r_0^6 \int\limits_{r \leq r_0} u^2 r^{1-n} dx .$$

Wählt man r_0 gemäß der Beziehung $4c_6 M r_0^2 \leq 1$, d. h. setzt man $r_1 = 1/2 (M c_6)^{-1/2}$, so folgt die behauptete Abschätzung mit

$$c_9 = 4c_6 , \quad c_{10} = 2c_6 M r_1^{10} + c_7 r_1^6 .$$

§ 4. Abschätzung von $\int u^2 r^{-1-n} dx$

Als nächstes benutzen wir den in § 3 bewiesenen Hilfssatz 4, um das Integral $\int u^2 r^{-1-n} dx$ abzuschätzen.

Satz 7: Es seien die Voraussetzungen von Satz 6 erfüllt. Dann gibt es ein $r_1 > 0$ und Konstanten c_{11}, c_{12}, welche nur von r_1, ε, p, n und M abhängen, d. h. für $0 < r_0 \leq r_1$ gilt

$$\int\limits_{r \leq r_0} u^2 r^{-1-n} dx \leq c_{11} \int\limits_{r \leq r_0} (L(u))^2 r^{3-n} (r_0^2 - r^2)^4 dx + c_{12} \int\limits_{r \leq r_0} u^2 r^{1-n} dx .$$

Beweis: Man wende Hilfssatz 4 auf die Funktion $y(t) = u(t x^0)$, $0 \leq t \leq 1$, mit festem x^0, $|x^0| = r_0$ an und integriere die so gewonnene Abschätzung über die Einheitskugel. Dann ergibt sich, wie früher beim Beweis von Satz 4, die Ungleichung

$$\int\limits_{r \leq r_0} u^2 r^{-1-n} dx \leq 4 r_0^2 \int\limits_{r \leq r_0} (\partial u/\partial r)^2 r^{1-n} (r_0^2 - r^2)^2 dx .$$

Folglich hat man

$$\int\limits_{r \leq r_0} u^2 r^{-1-n} dx \leq 4 r_0^2 \int\limits_{r \leq r_0} \sum_{i=1}^{n} (\partial u/\partial x_i)^2 r^{1-n} (r_0^2 - r^2)^2 dx \leq \\ \leq c_{13} \int\limits_{r \leq r_0} \sum_{i=1}^{n} (\partial u/\partial x_i)^2 r^{1-n} (r_0^2 - r^2)^4 dx + c_{14} \int\limits_{r \leq r_0} u^2 r^{1-n} dx ,$$

wobei c_{13}, c_{14} nur von n und r_0 abhängen. Wendet man jetzt Satz 6 an, so folgt sofort die Behauptung.

§ 5. Apriori-Abschätzung der Hölderkonstante von u

Wir schreiten endlich zum Beweis von Satz 1. Die Vervollständigung dieses Beweises läuft im wesentlichen analog zu dem Verfahren in [1], § 7 und 8. Wegen Hilfssatz 1 genügt es offenbar zu zeigen, daß unter den Voraussetzungen von Satz 1 für hinreichend kleines, jedoch nach Angabe von n, p, P, q, Q, M, ε und τ bereits fest bestimmbares δ der Ausdruck

$$\mu_\delta = \sup_{x^0 \in B} \int\limits_{\substack{|x-x^0| \leq \delta \\ x \in B}} (u(x) - u(x^0))^2 |x - x^0|^{-n-1} dx$$

eine Ungleichung der Form

$$\mu_\delta \leq c_{15} \|f\|^2 + c_{16} \|u\|^2 + c_{17} \|\varphi\|_2'$$

erfüllt mit Konstanten c_{15}, c_{16}, c_{17}, die nur von n, p, P, q, Q, M, ε und τ abhängen: Hilfssatz 1 liefert dann eine Abschätzung für $H_{1/2,\delta}$ und somit allgemein für $H_{\alpha,\delta}$, $\alpha \leq 1/2$. Die einzige Schwierigkeit liegt dabei noch in der Betrachtung der Randterme.

Um einen beliebigen Punkt x^0 des Randes Γ beschreiben wir eine Kugel $|x - x^0| \leq \mu$ (mit hinreichend kleinem Radius μ) und bilden durch eine zweimal stetig differenzierbare Transformation $x = x(y)$, oder ausgeschrieben

$$x_i = x_i(y_1, \ldots, y_n), \quad i = 1, \ldots, n,$$

der unabhängigen Variablen x mit nichtverschwindender Funktionaldeterminante den Durchschnitt dieser Kugel mit dem Bereich B topologisch so auf einen Bereich B^* des $(y_1, \ldots, y_n)$-Raumes ab, daß x^0 in den Nullpunkt übergeht und daß das Hyperflächenstück $|x - x^0| \leq \mu$, $x \in \Gamma$ in eine Umgebung von $y = 0$ auf der Hyperebene $y_1 = 0$ übergeht. Für $|x - x^0| \leq \mu$, $x \in B$, $x \notin \Gamma$ gelte ferner $y_1 > 0$.

Die Transformation $x = x(y)$ ist durch diese Eigenschaften natürlich nicht eindeutig bestimmt. Mit ihr leistet z. B. jede Abbildung der Form $x = x(Vy)$ genau dasselbe, wenn V eine beliebige nichtsinguläre konstante $n \times n$-Matrix von positiver Determinante ist, die die Ebene $y_1 = 0$ auf sich abbildet.

Sei die Matrix $((\partial y_k/\partial x_i))$ mit $S(x)$ bezeichnet, dann transformiert sich bei der Abbildung $x = x(y)$ der in Satz 1 definierte Operator L in den Operator

$$L^* = \sum_{i,k=1}^{n} a_{ik}^*(y)\, \partial^2/\partial y_i\, \partial y_k + \sum_{i=1}^{n} b_i^*(y)\, \partial/\partial y_i + b_0^*(y)$$

mit

$$((a_{ik}^*(y))) = S'(x(y))\, A(x(y))\, S((x(y)), \quad A(x) = ((a_{ik}(x)))$$

sowie

$$b_i^*(y) = \sum_{j,l=1}^{n} a_{jl}(x(y))\, \partial^2 y_i/\partial x_i\, \partial x_k + b_j(x(y))\, \partial y_i/\partial x_j, \quad b_0^*(y) = b_0(x(y)).$$

Man definiere nun $V = S'(x^0)\, T_{x^0}^{-1} O$ mit einer konstanten orthogonalen Matrix O, die dafür sorgt, daß V, wie verlangt, eine positive Determinante hat und die Ebene $y_1 = 0$ auf sich transformiert. Eine solche Matrix O kann immer

gefunden werden. T_{x^0} sei die in Satz 1 definierte nichtsinguläre Matrix. Wendet man anstatt der zuerst definierten Transformation die Transformation $x = x(Vy)$ an, so geht L über in den Operator

$$L^{**} = \sum_{i,k=1}^{n} a_{ik}^{**}(y)\, \partial^2/\partial y_i\, \partial y_k + \sum_{i=1}^{n} b_i^{**}(y)\, \partial/\partial y_i + b_0^{**}(y)$$

mit

$$\begin{aligned}((a_{ik}^{**}(y))) &= V^{-1} S'(x(Vy))\, A(x(Vy))\, S((x(Vy))\, V^{-1'} \\ &= O'\, T'_{x^0}\, S'^{-1}(x^0)\, S'(x(Vy))\, A(x(Vy))\, S(x(Vy))\, S^{-1}(x^0)\, T_{x^0} O\,.\end{aligned}$$

Für $y = 0$ ergibt sich also $A^{**}(0) = O'\, T'_{x^0} A(x^0)\, T_{x^0} O$, und diese Matrix genügt nach Voraussetzung der K_ε-Bedingung mit dem in den Voraussetzungen von Satz 1 angegebenen ε, denn die orthogonale Transformation O läßt die Eigenwerte der Matrix $A_{x^0}(x^0) = T'_{x^0} A(x^0)\, T_{x^0}$, also auch jede K_ε-Bedingung invariant. Wir wollen zeigen, daß für $A^{**}(y)$ in einer hinreichend kleinen Halbumgebung $y_1 \geqq 0$, $|y| < \tau'$ von $y = 0$ auch noch die $K_{\varepsilon/2}$-Bedingung erfüllt ist. Wesentlich ist dabei, daß τ' nicht von den speziellen Eigenschaften der $a_{ik}(x)$, sondern allein von der Auswahl der Transformation $x = x(y)$ (d. h. letzten Endes von der Gestalt des Bereiches B) und von den Konstanten n, p, P, q, Q, M, ε und τ aus Satz 1 abhängt.

Mit

$$Z(y) = ((z_{ik}(y))) = S(x(Vy))\, S^{-1}(x^0) - ((\delta_{ik}))$$

folgt

$$\begin{aligned}A^{**}(y) = O'\, T'_{x^0} A(x(Vy))\, T_{x^0} O\, + \\ + O'\, T'_{x^0} Z'(y)\, A(x(Vy))\, T_{x^0} O + O'\, T'_{x^0} A(x(Vy))\, Z(y)\, T_{x^0} O\, + \\ + O'\, T'_{x^0} Z'(y)\, A(x(Vy)) Z(y)\, T_{x^0} O\,.\end{aligned}$$

Wie bereits in der Einleitung ausgeführt, kann die K_ε-Bedingung auf die Komponenten der Matrix anstatt auf ihre Eigenwerte bezogen werden. Für die Matrix $A^{x^0} = ((a_{ik}^{x^0}))$ drückt sie sich dann aus als die Bedingung

$$(n-1) \sum_{i,k=1}^{n} (a_{ik}^{x^0})^2 \leqq (1 - \varepsilon/n) \left(\sum_{i=1}^{n} a_{ii}^{x^0} \right)^2,$$

und gemäß den Voraussetzungen von Satz 1 gilt dies für alle y mit $y_1 \geqq 0$ und $|y| \leqq \tau_1$ (τ_1 hinreichend klein). Daraus ergibt sich etwa für die Matrix

$$C = ((c_{ik})) = O'\, T'_{x^0} Z'(y)\, A(x(Vy))\, T_{x^0} O = O'\, T'_{x^0} Z'\, T'^{-1}_{x^0} A^{x^0}(x(Vy))\, O$$

die Abschätzung

$$\begin{aligned}\sum_{i,k=1}^{n} c_{ik}^2 \leqq \sum_{i,k=1}^{n} (t_{ik}^{x^0})^2 \sum_{i,k=1}^{n} z_{ik}^2 \sum_{i,k=1}^{n} (\hat{t}_{ik}^{x^0})^2 \sum_{i,k=1}^{n} (a_{ik}^{x^0}(x(Vy)))^2 \leqq \\ \leqq n^2 Q q^{-1} (1 - \varepsilon/n)/(n-1) \left(\sum_{i,k=1}^{n} z_{ik}^2(y) \right) \left(\sum_{i,k=1}^{n} a_{ii}^{x^0}(x(Vy)) \right)^2.\end{aligned}$$

Dabei sei $T_{x^0} = ((t_{ik}^{x^0}))$, $T_{x^0}^{-1} = ((\hat{t}_{ik}^{x^0}))$ gesetzt, q und Q seien die in den Voraussetzungen zu Satz 1 definierten unteren und oberen Schranken der Matrix T_{x^0}.

Ferner wurde zur Gewinnung der letzten Abschätzung mehrfach die Ungleichung

$$\sum_{i,k=1}^{n}\left(\sum_{l=1}^{n}\alpha_{il}\beta_{lk}\right)^{2}\leq\sum_{i,k=1}^{n}\alpha_{ik}^{2}\sum_{i,k=1}^{n}\beta_{ik}^{2}$$

benutzt, die für beliebige reelle Zahlen α_i, β_i, $i=1,\ldots,n$ gilt. Ganz entsprechend gewinnt man für die Quadratsummen über die Elemente der Matrizen

$$O'\,T'_{x^0}A(x(V\,y))(Z(y)\,T_{x^0}O$$

bzw.

$$O'\,T'_{x^0}Z'(y)A(x(V\,y))Z(y)\,T_{x^0}O$$

Abschätzungen durch Ausdrücke der Form

$$\sigma(n)\,Q\,q^{-1}\left(\sum_{i,k=1}^{n}z_{ik}^{2}(y)\right)\left(\sum_{i,k=1}^{n}a_{ii}^{x^0}(x(V\,y))\right)^{2}$$

bzw.

$$\sigma'(n)\,Q^{2}q^{-2}\left(\sum_{i,k=1}^{n}z_{ik}^{2}(y)\right)^{2}\left(\sum_{i,k=1}^{n}a_{ii}^{x^0}(x(V\,y))\right)^{2},$$

in welchen die Konstanten σ, σ' nur von n abhängen und die zusammen mit der zuerst hergeleiteten Ungleichung alle in einer gemeinsamen Umgebung $y_1\geq 0$, $|y|\leq\tau_1$ gelten. Somit erhält man insgesamt

$$\sum_{i,k=1}^{n}(a_{ik}^{**}(y))^{2}\leq[(1-\varepsilon/n)/(n-1)+\sigma''\,y^{2}]\left(\sum_{i=1}^{n}a_{ii}^{x^0}(x\,V\,y))\right)^{2},$$

denn es sind offenbar die Koeffizienten $z_{ik}(y)$ der Matrix $Z(y)$ stetig differenzierbar, da $x(y)$ zweimal stetig differenzierbar ist und die Funktionaldeterminante nach Voraussetzung nicht verschwindet, ferner weil

$$Z(0)=S(x^{0})\,S(x^{0})^{-1}-((\delta_{ik}))=0$$

gilt. Wesentlich ist wiederum, daß die Konstante σ'' nur von den Konstanten n, p, P, q, Q, M, ε und τ und von den speziellen Eigenschaften der Transformation $x(y)$, d. h. des Bereiches B abhängt. Als nächstes haben wir $\left(\sum_{i=1}^{n}a_{ii}^{x^0}(x(V\,y))\right)^{2}$ durch $\left(\sum_{i=1}^{n}a_{ii}^{**}(y)\right)^{2}$ abzuschätzen. Offenbar hat man dazu

$$|\mathrm{Spur}(A^{**})|\geq|\mathrm{Spur}(A^{x^0}(x(Vy)))|-\mathrm{Spur}\,[O\,A^{**}O'-A^{x^0}(x(V\,y))]\,.$$

Die Differenz $O\,A^{**}O'-A^{x^0}(x(V\,y))$ besteht wiederum aus den drei oben diskutierten Matrizen. Wegen $\left|\sum_{i=1}^{n}c_{ii}\right|^{2}\leq n\sum_{i,k=1}^{n}c_{ik}^{2}$ können dieselben Abschätzungen wie oben noch einmal benutzt werden, um eine Ungleichung der Form

$$(\mathrm{Spur}(A^{**}))^{2}\geq(1-\sigma'''\,|y|^{2})\,(\mathrm{Spur}(A^{x^0}(x(V\,y))))^{2}$$

herzuleiten, wobei σ''' dieselben Eigenschaften wie σ'' besitzt. Aus den beiden Ungleichungen zusammen folgt

$$\sum_{i,k=1}^{n}(a_{ik}^{**}(y))^{2}\leq[(1-\varepsilon/n)/(n-1)+\sigma''\,|y|^{2}]\,[1-\sigma'''\,|y|^{2}]^{-1}\left(\sum_{i=1}^{n}a_{ii}^{*}(y)\right)^{2}$$

für $y_1 \geq 0$, $|y| \leq \tau_2$. Es gilt

$$[(1-\varepsilon/n)/(n-1) + \sigma|y|^2]\,[1-\sigma'''|y|^2]^{-1} \leq (1-\varepsilon/(2n))/(n-1)$$

für $|y| \leq \tau_3$ bei hinreichend kleinem τ_3, das nur von ε, σ'', σ''' und n abhängt. Mit $\tau' = \mathrm{Min}(\tau_2, \tau_3)$ sind dann alle Voraussetzungen erfüllt, d. h. A^{**} genügt der $K_{\varepsilon/2}$-Bedingung für $y_1 \geq 0$, $|y| \leq \tau'$.

Wenn wir jetzt die Transformation $x = x(Vy)$ schlechthin wieder mit $x = x(y)$ bezeichnen, so können wir sagen: Die Transformation $x = x(y)$ kann so gewählt werden, daß sie den Bereich $|x - x^0| \leq \mu$, $x \in B$ zweimal stetig differenzierbar und umkehrbar eindeutig auf eine Umgebung des Nullpunktes $y = 0$ relativ zum Halbraum $y_1 \geq 0$ abbildet und derart, daß nach Transformation auf die Variablen y der Operator L die Gestalt

$$L^* = \sum_{i,k=1}^{n} a^*_{ik}(y)\, \partial^2/\partial y_i\, \partial y_k + \sum_{i=1}^{n} b^*_i(y)\, \partial/\partial y_i + b^*_0(y)$$

annimmt mit einer Matrix $A^*(y) = ((a^*_{ik}(y)))$, die für $|y| \leq \tau'$, $y_1 \geq 0$ der $K_{\varepsilon/2}$-Bedingung genügt, wogegen die Funktionen $b^*_i(y)$, $i = 0, 1, \ldots, n$, gleichmäßig beschränkt sind durch eine Konstante K, die nur von M und der speziellen Gestalt der Transformation $x(y)$ abhängt. Der Radius $\tau' = \tau'_{x^0}$ der obigen Halbkugel kann ebenfalls nach Vorgabe der Konstanten n, p, P, q, Q, M, ε und τ sowie der Transformation $x(y)$ bereits bestimmt werden.

Offenbar gibt es nun eine Kugel $\mathfrak{K}_{x^0} = \{x \ni |x - x^0| \leq r_{x^0}\}$, so daß der Durchschnitt $\mathfrak{K}_{x^0} \cap B$ mit dem Durchschnitt von $\mathfrak{K}_{x^0}$ und der Bildmenge im x-Raum der Halbkugel $|y| \leq \tau'_{x^0}$, $y_1 \geq 0$ übereinstimmt. Wir bezeichnen den Durchschnitt von B und der Kugel $|x - x^0| \leq r_{x^0}/2$ mit F_{x^0}. Das Bild $y(F_{x^0})$ des Bereiches F_{x^0} im y-Raum hat offenbar einen positiven Abstand s_{x^0} von der Kugelfläche $|y| = \tau'_{x^0}$. Wenn daher x' irgendeinen Punkt aus F_{x^0} bezeichnet und y' sein Bild, so liegt die Kugel $|y - y'| \leq s_{x^0}$ stets innerhalb der Kugel $|y| \leq \tau'_{x^0}$. Ferner gibt es eine positive Zahl s'_{x^0} derart, daß das y-Bild des Durchschnitts der Kugel $|x - x'| \leq s'_{x^0}$ mit dem Bereich B innerhalb von $|y - y'| \leq s_{x^0}$ liegt, und s'_{x^0} kann unabhängig von der Wahl des Punktes x' im Bereich F_{x^0} gewählt werden. Offenbar hängen wiederum s'_{x^0} und s_{x^0} allein von der Transformation $x(y)$ und von der Zahl τ_{x^0} ab. Für irgendein $u(x)$ mit stückweise stetigen zweiten partiellen Ableitungen in B und für beliebiges x' aus F_{x^0} hat man nun

$$\int\limits_{x \in B,\, |x - x'| \leq s'_{x^0}} (u(x) - u(x'))^2 |x - x'|^{-n-1} dx \leq$$

$$\leq c_{18} \int\limits_{y_1 \geq 0,\, |y - y'| \leq s_{x^0}} (v(y) - v(y'))^2 |y - y'|^{-n-1} dy\,,$$

wo $v(y) = u(x(y))$ und c_{18} eine gewisse, nur von s_{x^0} und den Eigenschaften der Transformation $x(y)$ abhängige Konstante ist. (Insbesondere ist c_{18} unabhängig von $u(x)$.)

Wenn $u(x)$ eine Lösung von $L(u) = f$ ist, so gilt

$$L^*(v) = f(x(y)) = g(y)\,.$$

Für hinreichend klein gewähltes s_{x^0} erfüllt L^* die Voraussetzungen von Satz 7 mit den Konstanten s'_{x^0}, $\varepsilon/2$ statt r_1, ε und mit gewissen durch die Transformation $x(y)$ bestimmten neuen Zahlen p^* und M^* statt p und M, allerdings nur im Bereich $|y-y'| \leqq s_{x^0}$, $y_1 \geqq 0$. Wenn $u(x) = \varphi(x)$ für $x \in \Gamma$, so folgt $v(y) = \varphi(x(y)) = \psi(y)$ für $y_1 = 0$. Setzen wir die Funktion $\psi(y)$ für $y_1 \neq 0$ durch die Bedingung

$$\psi(y_1, \ldots, y_n) = \psi(0, y_2, \ldots, y_n)$$

fort, so gilt offenbar

$$\operatorname*{Max}_{i,k=1, |y| \leqq \tau'_{x^0}}^{n} \{|\psi|, |\partial\psi/\partial y_i|, |\partial^2\psi/\partial y_i\, \partial y_k|\} \leqq c_{19} \|\varphi\|'_2 .$$

Sei $w = v - \psi$, so folgt $L^*(w) = h$ mit $h(y) = g(y) - L^*(\psi)$; $w = 0$ für $y_1 = 0$, und es gilt offenbar

$$\operatorname*{Max}_{y_1 \geqq 0, |y| \leqq \tau'_{x^0}} |h| \leqq c_{20} \|\varphi\|'_2 + c_{21} \|f\| \cdot$$

Jetzt werde an der Ebene $y_1 = 0$ gespiegelt, und zwar sollen $w(y)$ und $h(y)$ ungerade fortgesetzt werden. Dann ist $w(y)$ stetig differenzierbar und hat stückweise stetige zweite Ableitungen in der ganzen Kugel $|y| \leqq \tau'_{x^0}$. Setzt man auch a^*_{ik}, b^*_i, $i = 0, \ldots, n$, fort, und zwar b^*_0, a^*_{11}, $a^*_{\lambda\mu}$, b^*_λ, $\lambda, \mu = 2, \ldots, n$ gerade und b^*_1, $a^*_{\lambda 1}$ $a^*_{1\lambda}$, $\lambda = 2, \ldots, n$, ungerade, so gilt auch in der ganzen Kugel $|y| \leqq \tau'_{x^0}$ stets $L^*(w) = h$. Alle Voraussetzungen von Satz 7 bleiben auch in der vollen Kugel $|y-y'| \leqq s_{x^0}$ erfüllt. Folglich liefert Satz 7:

$$(*)\left\{\begin{aligned}
&\int\limits_{x \in B, |x-x'| \leqq s'_{x^0}} (u(x) - u(x'))^2 |x-x'|^{-n-1} dx \leqq \\
&\leqq c_{18} \int\limits_{y_1 \geqq 0, |y-y'| \leqq s_{x^0}} (v(y) - v(y'))^2 |y-y'|^{-n-1} dy \leqq \\
&\leqq 2c_{18} \int\limits_{|y-y'| \leqq s_{x^0}} |w(y) - w(y')|^2 |y-y'|^{-n-1} dy + \\
&+ 2c_{18} \int\limits_{|y-y'| \leqq s_{x^0}} (\psi(y) - \psi(y'))^2 |y-y'|^{-n-1} dy \leqq \\
&\leqq 2c_{11}c_{18} \int\limits_{|y-y'| \leqq s_{x^0}} [L^*(w(y) - w(y'))]^2 |y-y'|^{3-n} (s^2_{x^0} - |y-y'|^2)^4 dy + \\
&+ 2c_{12}c_{18} \int\limits_{|y-y'| \leqq s_{x^0}} (w(y) - w(y'))^2 |y-y'|^{1-n} dy + \\
&+ 2c_{18} \int\limits_{|y-y'| \leqq s_{x^0}} (\psi(y) - \psi(y'))^2 |y-y'|^{-n-1} dy \leqq \\
&\leqq c_{22} \|\varphi\|'^2_2 + c_{23} \|f\|^2 + c_{24} \|u\|^2 .
\end{aligned}\right.$$

Letzteres folgt, weil einerseits $L^*(w) = h$ wie oben abgeschätzt werden kann und weil man andererseits hat $L^*(w(y')) = b^*_0(y)\, w(y')$, ferner weil die beiden ersten Integrale eine Potenz von $|y-y'|$ enthalten, die eine Abschätzung durch das Maximum des absoluten Betrages erlauben, während für das letzte Integral eine Abschätzung durch die Ableitungen von ψ möglich ist.

Sei jetzt x^0 ein Punkt aus dem Innern von B und $r_{x^0} > 0$ so beschaffen, daß $r_{x^0} < \tau$ gilt und daß die Kugel $\mathfrak{K}_{x^0} = \{x \ni |x - x^0| \leq r_{x^0}\}$ noch ganz innerhalb von B liegt. Hierbei sei τ die in Satz 1 definierte Konstante. Innerhalb von $\mathfrak{K}_{x^0}$ führen wir neue Koordinaten $y = (y_1, \ldots, y_n)$ durch die lineare Transformation $x - x^0 = T_{x^0}^{-1} y$ ein. Dann geht L in den Differentialausdruck

$$L^* = \sum_{i,k=1}^{n} a_{ik}^* \partial^2/\partial y_i \, \partial y_k + \sum_{i=1}^{n} b_i^* \partial/\partial y_i + b_0^*$$

über mit

$$A^* = ((a_{ik}^*)) = T_{x^0} A \, T_{x^0}' \quad \text{und} \quad b_i^* = \sum_{k=1}^{n} t_{ik} b_k, \; i = 1, \ldots, n, \; b_0^* = b_0 \, .$$

Sei x' ein Punkt aus B mit $|x' - x^0| \leq r_{x^0}/2$ und sei

$$\mathfrak{K}' = \{x \ni |x - x'| \leq 1/2 \, r_{x_0} q/Q\} \, .$$

Die Abbilder von $\mathfrak{K}_{x^0}$ und $\mathfrak{K}'$ unter obiger Transformation seien mit $\mathfrak{E}_{x^0}$ und $\mathfrak{E}'$ bezeichnet. Dann hat man

$$\begin{aligned} \mathfrak{E}_{x^0} &= \{y \ni |T_{x^0} y| \leq r_{x^0}\} \\ \mathfrak{E}' &= \{y \ni |T_{x^0}(y - y')| \leq 1/2 \, r_{x^0} q/Q\} \, , \end{aligned}$$

wobei $y' = T_{x^0}^{-1}(x' - x^0)$ das Bild des Punktes x' im y-Raum bezeichne.

Offenbar gilt $\mathfrak{K}' \subset \mathfrak{K}_{x^0}$ und daher auch $\mathfrak{E}' \subset \mathfrak{E}_{x^0}$. Mehr noch, das Ellipsoid $\mathfrak{E}'$ ist in der Kugel $|y - y'| \leq 1/2 \, r_{x^0}/Q$ enthalten, welche ihrerseits ein Teilbereich von $\mathfrak{E}_{x^0}$ ist. Um die erste Inklusion einzusehen, beachte man, daß einerseits für $y \in \mathfrak{E}'$ die Ungleichung

$$q \, |y - y'| \leq |T_{x^0}(y - y')| \leq 1/2 \, r_{x^0} q/Q \, ,$$

also die Abschätzung

$$|y - y'| \leq 1/2 \, r_{x^0}/Q$$

erfüllt ist. Für die zweite Inklusion bemerken wir andererseits, daß die Ungleichung $|y - y'| \leq 1/2 \, r_{x^0}/Q$ die Abschätzung

$$|x - x'| = |T_{x^0}(y - y')| \leq Q \, |y - y'| \leq 1/2 \, r_{x^0}$$

nach sich zieht. Da $|x' - x^0| \leq r_0/2$ gilt, folgt also

$$|x - x^0| \leq |x - x'| + |x - x^0| \leq r_{x^0}$$

und somit $x \in \mathfrak{K}_{x^0}$, $y \in \mathfrak{E}_{x^0}$. Daraus leitet man her, daß

$$\begin{aligned} &\int\limits_{|x - x'| \leq 1/2 r_{x^0} q/Q} (u(x) - u(x'))^2 \, |x - x'|^{-n-1} dx \leq \\ &\leq Q^n/q^{n+1} \int\limits_{|y - y'| \leq 1/2 r_{x^0}/Q} (v(y) - v(y'))^2 \, |y - y'|^{-n-1} dy \end{aligned}$$

gilt für jedes in $|x - x^0| \leq r_{x^0}$ zweimal stetig differenzierbare $u(x)$ und mit $v(y) = u(x^0 + T_{x^0}^{-1} y)$. Wenn nun $u(x)$ eine Lösung von $L(u) = f$ ist, so gilt

$$L^*(v) = f(x^0 + T_{x^0}^{-1} y) \, .$$

Für hinreichend klein gewähltes $r_{x'}$ erfüllt der Operator L^* in der Kugel $|y - y'| \leq 1/2\, r_{x'}/Q$ ferner alle Voraussetzungen von Satz 7 mit den Konstanten $r_1 = 1/2\, r_{x'}/Q$, ε, n und mit $p^* = q^2 p$ sowie $M^* = (Q^2 + 1)M$ anstatt p und M beziehungsweise.

Folglich hat man auf ähnliche Weise wie für die Randterme:

$$(**)\left\{\int\limits_{|x - x'| \leq {}^1/_2 r_{x'} q/Q} (u(x) - u(x'))^2\, |x - x'|^{-n-1} \leq c_{25}\, \|u\|^2 + c_{26}\, \|f\|^2 .\right.$$

Endlich ergibt sich der Beweis von Satz 1 folgendermaßen: Nach dem Heine-Borelschen Überdeckungssatz ist es möglich, eine endliche Anzahl von Punkten x^ν, $\nu = 1, \ldots, N_1$ auf dem Rand Γ und eine endliche Anzahl von Punkten $x^{\nu'}$, $\nu = 1, \ldots, N_2$ im Innern von B zu finden, derart, daß die Bereiche F_{x^ν}, $\nu = 1, \ldots, N_1$ und die Kugeln $|x - x^\nu| \leq r_{x^{\nu'}}/2$ den Bereich B vollständig überdecken. Wenn man dann mit δ das Minimum der endlich vielen positiven Zahlen s_{x^ν}, $\nu = 1, \ldots, N_1$ und $1/2\, r_{x^{\nu'}}/Q$, $\nu = 1, \ldots, N_2$ bezeichnet, so ergibt sich offenbar

$$\int\limits_{x \in B,\, |x - x'| \leq \delta} (u(x) - u(x'))^2\, |x - x'|^{-n-1} d x \leq c_{15}\, \|f\|^2 + c_{16}\, \|u\|^2 + c_{17}\, \|\varphi\|_2'^2$$

mit gewissen Konstanten c_{15}, c_{16}, c_{17}, die allein von n, p, P, q, Q, ε und τ abhängen. Denn für beliebiges $x' \in B$ gilt entweder $x' \in F_{x^\nu}$ für ein gewisses ν oder $|x' - x^{\nu'}| \leq 1/2\, r_{x^{\nu'}}\, q/Q$ für ein gewisses ν'. Im ersten Falle schätzt man auf Grund von Formel (*), im zweiten Falle auf Grund von Formel (**) ab. Damit folgt nach Hilfssatz 1 der Beweis von Satz 1 unmittelbar. Man kann sogar immer den Hölderexponenten $\alpha = 1/2$ benutzen, auf jeden Fall aber jedes beliebige α zwischen Null und 1/2.

Literatur

[1] Cordes, H. O.: Über die erste Randwertaufgabe bei quasilinearen Differentialgleichungen zweiter Ordnung in mehr als zwei Variablen p. Math. Ann. **131**, 278—312 (1956). — [2] Leray, J., and I. Schauder: Topologie et équations fonctionelles. Ann. Sci. Ec. norm. sup. **51**, 45—78 (1934). — [3] Morrey, C. B.: On the solutions of quasilinear elliptic partial differential equations. Trans. Amer. math. Soc. **43**, 126—166 (1938). — [4] Nirenberg, L.: On nonlinear elliptic partial differential equations and Hölder-continuity. Commun. pure appl. Math. **6**, 103—156 (1953).

(Eingegangen am 19. Februar 1959)

Jörgens, K.
Math. Annalen 138, 179—202 (1959)

Über die nichtlinearen Wellengleichungen der mathematischen Physik*)

Von
Konrad Jörgens in Heidelberg

Einleitung

Die nichtlinearen partiellen Differentialgleichungen, die sich formal aus einem Variationsprinzip mit einer Lagrangefunktion der Form

$$L = L(K, I)\,, \quad K = u_x^2 + u_y^2 + u_z^2 - u_t^2\,, \quad I = u^2$$

herleiten lassen, sind als Verallgemeinerungen der linearen Wellengleichung

$$u_{tt} - u_{xx} - u_{yy} - u_{zz} + \mu^2 u = 0$$

aufzufassen, die aus $L = \frac{1}{2}(K + \mu^2 I)$ hergeleitet ist. Die Herkunft aus einem Variationsprinzip und die Invarianz der Lagrangefunktion gegenüber den inhomogenen Lorentztransformationen garantieren die Gültigkeit der zehn Erhaltungssätze. Eine solche Gleichung soll als *nichtlineare Wellengleichung* bezeichnet werden, wenn sie auch noch die folgenden Eigenschaften mit der linearen Wellengleichung gemein hat:

1) Die Differentialgleichung ist *hyperbolisch*. Das Anfangswertproblem in bezug auf eine raumartige Anfangsmannigfaltigkeit ist demnach sachgemäß gestellt.

2) Die Charakteristiken sind *zeitartig*, d. h. das Einflußgebiet jedes Punktes P ist enthalten in der Zukunfts-Hälfte des Lichtkegels mit der Spitze in P, das Abhängigkeitsgebiet in der Vergangenheits-Hälfte (Kausalität).

3) Das Energie-Integral ist *positiv* für alle Funktionen $u \not\equiv 0$, falls es existiert, und ist gleich Null für $u \equiv 0$.

Diese drei Forderungen lassen sich leicht als Ungleichungen formulieren, denen die Lagrangefunktion und deren Ableitungen genügen müssen.

Nichtlineare Wellengleichungen kommen in der Mesonenphysik vor, da sie geeignet sind, Wellenfelder mit Selbst-Wechselwirkung zu beschreiben; und zwar sind bisher zwei Typen von Gleichungen betrachtet worden: Die *semilinearen* Gln. mit der Lagrangefunktion $L = \frac{1}{2}(K + F(I))$ und die Gln. vom Bornschen Typ mit $L = \sqrt{K + F(I)} - G(I)$[1]. Zum ersten Typ gehört die

*) Bei der Naturwissenschaftlich-Mathematischen Fakultät der Universität Heidelberg als Habilitationsschrift eingereicht.

[1]) M. Born hat 1934 (vgl. [3]) nichtlineare Gln. des elektromagnetischen Feldes vorgeschlagen, deren Lagrangefunktion die Quadratwurzel einer Lorentzinvarianten des Feldes ist. Die Idee einer skalaren Theorie dieser Art stammt von Heisenberg (1939).

Mesonengleichung von SCHIFF [11], zum zweiten die von HEISENBERG [6]. Gerade diese Beispiele legen es nun nahe, eine zusätzliche Bedingung an die Wellengleichung zu stellen: Soll nämlich eine Wellengl. die zeitliche Veränderung eines Mesonen-Wellenfeldes beschreiben, so muß zu *beliebigen*, physikalisch zulässigen Anfangswerten u, u_t, die zur Zeit $t = 0$ vorgeschrieben sind, genau eine Lösung der Wellengl. im ganzen Halbraum $t \geqq 0$ existieren, die für $t = 0$ die vorgeschriebenen Werte annimmt[2]). Ist dies der Fall, so sagt man, das Anfangswertproblem sei *im Großen* lösbar. Als physikalisch zulässig bezeichnet man Anfangswerte u, u_t, die hinreichend oft differenzierbar und so beschaffen sind, daß das Energie-Integral konvergiert. Hinsichtlich der Lösung u des Anfangswertproblems hat man zwischen *eigentlichen*, d. h. mindestens zweimal stetig differenzierbaren, und *schwachen* Lösungen (Stoßwellen) zu unterscheiden. Das Problem ist, alle Wellengleichungen zu bestimmen, für die das Anfangswertproblem, sei es im eigentlichen, sei es im schwachen Sinne, im großen lösbar ist.

In dieser Arbeit wird folgendes gezeigt: Es existiert eine *ausgezeichnete* Klasse von Wellengleichungen, die aus den semilinearen und den Bornschen Gln. besteht sowie aus allen den Gln., die man aus diesen durch Transformation der abhängigen Variablen erhält. Für alle Wellengln., die nicht zu dieser Klasse gehören, ist das Anfangswertproblem im großen im eigentlichen Sinne *nicht* lösbar, d. h. die Lösung bleibt im allgemeinen nur innerhalb eines endlichen Zeitintervalls zweimal stetig differenzierbar; wenn jedoch schwache Lösungen des Anfangswertproblems existieren, so genügen sie im allgemeinen nicht den Erhaltungssätzen der Energie und des Impulses, scheiden also für die Beschreibung physikalischer Vorgänge aus. Für die schwachen Lösungen der ausgezeichneten Wellengln. sind die Erhaltungssätze hingegen stets erfüllt. Außerdem haben die ausgezeichneten Wellengln., und nur diese, die Eigenschaft, daß die Unstetigkeitsflächen der schwachen Lösungen Charakteristiken sind; die Sprunghöhe (der ersten Ableitungen beim Durchgang durch die Fläche) genügt einer gewöhnlichen Differentialgleichung längs der Bicharakteristiken, aus der folgt, daß die Sprunghöhe entweder nirgends oder überall auf einer Bicharakteristik verschwindet, d. h. daß die Unstetigkeiten der Lösung nur von Unstetigkeiten der Anfangswerte herrühren können. Diese Resultate gelten, mit einer Ausnahme, auch für Wellengln. für eine komplexwertige Funktion.

Sehr ähnlich liegen die Verhältnisse bei den nichtlinearen Theorien des elektromagnetischen Feldes. Hier nimmt die Bornsche Elektrodynamik eine Sonderstellung ein, wie von D. I. BLOCHINZEW und W. ORLOW [2] gezeigt worden ist. Insbesondere können in der Bornschen Elektrodynamik keine Stoßwellen entstehen. Dasselbe Resultat für eine spezielle skalare Bornsche Wellengleichung hat T. TANIUTI [12] gefunden.

Die gestellte Aufgabe wäre vollständig gelöst, wenn bewiesen würde, daß für die ausgezeichneten Wellengln. das Anfangswertproblem im großen lösbar ist. In dieser Arbeit wird nur die Existenz von (eigentlichen) Lösungen $u(x, t)$

[2]) Die Lösung existiert dann auch für $t < 0$.

bewiesen, die von zwei Variablen abhängen, und zwar bei den semilinearen Gln. ohne wesentliche weitere Einschränkung und auch für komplexwertige Lösungen; bei den Bornschen Gln. aber nur im Fall $G(I) = \text{const}$ und in charakteristischen Variablen. Diese Lösungen der Bornschen Gln. sind als Funktionen von x und t zwar in der ganzen Halbebene $t \geqq 0$ definiert, jedoch unter Umständen mehrdeutig. Es ist klar, daß nur eindeutige Lösungen physikalische Bedeutung haben; es scheint aber nicht leicht zu sein, Bedingungen für die Eindeutigkeit anzugeben.

Teile dieser Arbeit sind entstanden, während ich Mitarbeiter im Max-Planck-Institut für Physik in Göttingen (jetzt München) bzw. Gast im Institute of Mathematical Sciences der New York University war. Es ist mir eine angenehme Pflicht, den Herren Prof. W. Heisenberg und Prof. A. Schlüter vom Max-Planck-Institut für Physik sowie Prof. P. D. Lax und Prof. G. B. Whitham vom Institute of Mathematical Sciences für nützliche Anregungen und für ihr Interesse an dieser Arbeit herzlich zu danken.

§ 1. Wellengleichungen

Es werden die folgenden Bezeichnungen benutzt:

$$x = \{x^0, x^1, x^2, x^3\}, \quad (x^0 = t)$$

$$g^{ij} = \begin{cases} -1 & \text{für} \quad i = j = 0 \\ 1 & \text{für} \quad i = j = 1, 2, 3 \\ 0 & \text{für} \quad i \neq j \end{cases}$$

$$u_i = \frac{\partial u}{\partial x^i}, \quad u_{ij} = \frac{\partial^2 u}{\partial x^i \partial x^j}, \quad u^i = g^{ij} u_j \left(= \sum_{j=0}^{3} g^{ij} u_j \right)$$

$$I = u^2, \quad K = K\{u\} = g^{ij} u_i u_j = u^j u_j \left(= \sum_{j=0}^{3} u^j u_j \right),$$

wobei das Summenzeichen stets weggelassen wird.

Die Lagrangefunktion $L = L(K, I)$ sei in einem Teilgebiet $\mathfrak{G}$ der KI-Ebene erklärt und fünfmal stetig differenzierbar. Das Gebiet $\mathfrak{G}$ habe die Form

$$\mathfrak{G}: \begin{cases} 0 \leqq I < \infty \\ K_0(I) < K < \infty, \; K_0(I) < 0 \quad \text{für} \quad 0 \leqq I < \infty. \end{cases} \tag{1.1}$$

Bei den semilinearen Gln. ist $K_0(I) = -\infty$, bei den Bornschen Gln. ist $K_0(I) = -F(I)$. $K_0(I)$ wird als negativ angenommen, damit die Lagrangefunktion für alle ebenen Wellen, deren Geschwindigkeit nicht größer als die Lichtgeschwindigkeit ist, d. h. für alle Funktionen $u = f(\xi_i x^i)$ mit $\xi^i \xi_i \geqq 0$, erklärt ist. Für diese Funktionen ist nämlich $K\{u\} = f'^2 \xi^i \xi_i \geqq 0$. Die zu L gehörige Eulersche Gleichung ist

$$-\frac{\partial}{\partial x^i}\{L_K u^i\} + L_I u = 0 \tag{1.2}$$

oder

$$\{L_K g^{ij} + 2 L_{KK} u^i u^j\} u_{ij} + \{2 L_{KI} K - L_I\} u = 0. \tag{1.2a}$$

Setzt man

$$T^{ij} = 2 L_K u^i u^j - L g^{ij}, \tag{1.3}$$

so folgen aus (1.2) die Erhaltungssätze der Energie und des Impulses[3])

$$\frac{\partial T^{ij}}{\partial x^j} = 2u^i\left\{\frac{\partial}{\partial x^j}(L_K u^j) - L_I u\right\} = 0 . \tag{1.4}$$

T^{00} ist die Energiedichte und $\{T^{01}, T^{02}, T^{03}\}$ der Impulsdichtevektor. Aus (1.4) folgt die Integralform der Erhaltungssätze

$$\int_{\mathfrak{F}} T^{ij} n_j \, dS = 0 ; \tag{1.4a}$$

$\mathfrak{F}$ ist eine beliebige geschlossene Fläche, n_i der Normalvektor $\left(\text{mit } \sum_{i=0}^{3} (n_i)^2 = 1\right)$ und dS das Oberflächenelement.

Die Differentialgleichung (1.2) ist eine Wellengleichung, wenn die folgenden drei Bedingungen erfüllt sind:

1) Die Differentialgleichung ist hyperbolisch; eine notwendige Bedingung hierfür ist, daß die Koeffizientenmatrix des Hauptteils

$$A^{ij} = L_K g^{ij} + 2L_{KK} u^i u^j$$

eine negative Determinante hat. Man erhält

$$\det(A^{ij}) = -(L_K)^3 \{L_K + 2KL_{KK}\} < 0 .$$

Daraus folgt $L_K \neq 0$, also ohne Beschränkung der Allgemeinheit

$$L_K > 0 , \quad L_K + 2KL_{KK} > 0 . \tag{1.5}$$

Diese Ungleichungen sind aber auch hinreichend für den hyperbolischen Charakter der Differentialgleichung.

2) Die Charakteristiken sind zeitartig, d. h. ihr Normalvektor n_i genügt der Bedingung $n^i n_i \geqq 0$[4]). Die Charakteristiken sind bekanntlich durch die Gleichung

$$A^{ij} n_i n_j = L_K n^i n_i + 2L_{KK}(u^i n_i)^2 = 0$$

definiert. Die Bedingung $n^i n_i \geqq 0$ ist also wegen (1.5) der Ungleichung

$$L_{KK} \leqq 0 \tag{1.6}$$

äquivalent.

3) Das Energie-Integral

$$\iiint_{-\infty}^{+\infty} T^{00} \, dx^1 \, dx^2 \, dx^3$$

ist positiv für $u(x) \not\equiv 0$, falls es existiert, und gleich Null für $u(x) \equiv 0$. Hieraus folgen die Ungleichungen $L - 2KL_K \geqq 0$ für $K \leqq 0$, $L \geqq 0$ für $K > 0$, die jedoch nicht hinreichend sind und auch im folgenden nicht gebraucht werden. Für die ausgezeichneten Wellengln. werden notwendige und hinreichende Bedingungen in § 5 bzw. in § 6 aufgestellt.

Zur Charakterisierung der ausgezeichneten Klasse von Wellengln., die aus den semilinearen und den Bornschen Gln. sowie aus den daraus durch Trans-

[3]) Die übrigen Erhaltungssätze folgen aus diesen und aus der Symmetrie des Tensors T^{ij}.

[4]) Vgl. D. I. Blochinzew [1].

formation der abhängigen Variablen hervorgehenden besteht, wird die Funktion

$$M(K, I) = \{L_K(K, I)\}^{-2}$$

eingeführt. Die Koeffizientenmatrix des Hauptteils läßt sich durch M ausdrücken

$$A^{ij} = M^{-3/2}\{M g^{ij} - M_K u^i u^j\} \tag{1.7}$$

und die Ungleichungen (1.5), (1.6) gehen über in

$$M - K M_K > 0\,, \quad M_K \geqq 0\,. \tag{1.8}$$

Die ausgezeichnete Klasse von Wellengleichungen ist durch die Gleichung $M_{KK} \equiv 0$ *charakterisiert.* Aus dieser Gl. folgt nämlich $M_K = A(I)$, und es ist $A(I) \geqq 0$ nach (1.8). Ist $A(I) \equiv 0$, so erhält man $L = \frac{1}{2} B(I)K + C(I)$ mit $B(I) > 0$ wegen (1.5); ist aber $A(I) \not\equiv 0$, so folgt $L = \sqrt{B(I)\,K + C(I)} + D(I)$ mit $B(I) = 4A^{-1}$, und $B(I)$ ist positiv, weil sonst L für $K \neq 0$ und gewisse Werte von I nicht stetig wäre. Die Transformation

$$U = \int_0^u \sqrt{B(v^2)}\, dv$$

liefert dann im ersten Fall die Lagrangefunktion einer semilinearen Wellengleichung

$$L = \frac{1}{2}(K + F(I)) \tag{1.9}$$

und im zweiten Fall eine Lagrangefunktion vom Bornschen Typ

$$L = \sqrt{K + F(I)} - G(I) \quad \text{mit} \quad F(I) > 0\,, \tag{1.10}$$

und zwar ist $F(I) > 0$ wegen (1.1). Offenbar genügen diese Funktionen den Ungleichungen (1.5) und (1.6).

§ 2. Eigentliche Lösungen

In diesem Paragraph sollen zweimal stetig differenzierbare Lösungen einer Wellengleichung untersucht werden mit dem Ziel, den folgenden Satz zu beweisen:

Satz 2. *Die Wellengleichung gehöre nicht zur ausgezeichneten Klasse, d. h. es sei* $M_{KK} \not\equiv 0$ *im Definitionsgebiet* $\mathfrak{S}$ *der Lagrangefunktion* L. *Dann gibt es eine Lösung* $u(x, t)$ *der Wellengleichung, definiert in einem an die Gerade* $t = 0$ *anschließenden Teilgebiet der Halbebene* $t \geqq 0$, *mit den Eigenschaften:*

1) u *ist zweimal stetig differenzierbar in einem Gebiet der Form* $|x| \leqq 2t_0 - t$, $0 \leqq t < t_0$, *jedoch mindestens eine der zweiten Ableitungen von* u *wächst unbeschränkt bei Annäherung an einen Randpunkt mit* $t = t_0$.

2) *Die Anfangswerte* $u(x, 0)$, $u_t(x, 0)$ *sind unendlich oft differenzierbar.*

3) *Die Werte der Funktionen* $I = u^2$ *und* $K = K\{u\}$ *verbleiben für* $|x| \leqq 2t_0 - t$, $0 \leqq t < t_0$ *in einem abgeschlossenen Teilgebiet* $\mathfrak{S}'$ *des Definitionsgebietes* $\mathfrak{S}$ *der Lagrangefunktion.*

Der Satz besagt, daß das Anfangswertproblem im großen für eine solche Gleichung im eigentlichen Sinne nicht lösbar ist. Die Beschränkung auf zwei Variable ist dabei offenbar unwesentlich. Punkt 2) der Behauptung gewährleistet, daß die Singularität der Lösung nicht von einer Singularität der Anfangswerte herrührt, Punkt 3), daß die Singularität auch nicht durch Annäherung an den Rand des Regularitätsgebietes der Differentialgleichung entsteht. Ein entsprechender Satz für hyperbolische Systeme von zwei homogenen partiellen Differentialgleichungen erster Ordnung ist von P. D. LAX [8] bewiesen worden. Der folgende Beweis ist eine Verallgemeinerung der Methode von LAX.

Beweis. 1) Es wird $u_t = v$, $u_x = w$ gesetzt. Die Dgl. für $u(x, t)$ ist von der Form

$$a u_{tt} + 2b u_{tx} + c u_{xx} + d = 0 .$$

Die Koeffizienten hängen von u, v und w ab und sind in einem Gebiet $\mathfrak{H}$ des uvw-Raumes dreimal stetig differenzierbar. Nach (1.7) und (1.8) ist $a = M + v^2 M_K > 0$; die charakteristischen Geschwindigkeiten λ und μ sind gegeben durch

$$\left.\begin{matrix}\lambda\\ \mu\end{matrix}\right\} = \frac{-vw M_K \pm \sqrt{M^2 - K M M_K}}{M + v^2 M_K}$$

und genügen den Ungleichungen

$$-1 \leqq \mu < \lambda \leqq 1 .$$

Man berechnet

$$\left.\begin{matrix}\lambda\lambda_v - \lambda_w\\ \mu\mu_v - \mu_w\end{matrix}\right\} = \frac{M_{KK}}{\sqrt{M^2 - K M M_K}} \left\{\frac{wM \pm v\sqrt{M^2 - K M M_K}}{M + v^2 M_K}\right\}^3 .$$

Nach Voraussetzung sind diese Ausdrücke nicht beide identisch gleich Null in $\mathfrak{H}$. Es sei z. B. $\lambda\lambda_v - \lambda_w < 0$ in einem Teilgebiet $\mathfrak{H}'$ von $\mathfrak{H}$. Die Differentialgleichung ist dem System

$$\begin{aligned} u_t &= v \\ v_t + \lambda v_x + \mu(w_t + \lambda w_x) &= -\frac{d}{a} \\ v_t + \mu v_x + \lambda(w_t + \mu w_x) &= -\frac{d}{a} \end{aligned} \tag{2.1}$$

äquivalent. Man wähle zwei feste Lösungen $V(u, v, w)$, $W(u, v, w)$ der Dgln.

$$V_w = \mu V_v \quad \text{bzw.} \quad W_w = \lambda W_v ,$$

welche in einem Teilgebiet $\mathfrak{H}''$ von $\mathfrak{H}'$ dreimal stetig differenzierbar sind und dort den Ungleichungen $V_v > 0$, $W_v > 0$ genügen. Wegen

$$\frac{\partial(u, V, W)}{\partial(u, v, w)} = (\lambda - \mu)\, V_v W_v > 0$$

kann $\mathfrak{H}''$ insbesondere so gewählt werden, daß die Transformation

$$\mathfrak{H}'' \ni (u, v, w) \to (u, V, W) \in \mathfrak{K}$$

eineindeutig ist. Das System (2.1) wird durch diese Transformation in das

System

$$(2.2)\qquad \begin{aligned} u_t &= X \\ V_t + \lambda V_x &= Y \\ W_t + \mu W_x &= Z \end{aligned}$$

übergeführt, worin X, Y und Z zweimal stetig differenzierbare Funktionen von u, V und W sind. Man berechnet

$$\lambda_V = \frac{\lambda\lambda_v - \lambda_u}{V_v(\lambda - \mu)} < 0 \quad \text{in } \mathfrak{R}.$$

Mit den Bezeichnungen $\Phi = V_x$ und $\frac{d}{dt} = \frac{\partial}{\partial t} + \lambda \frac{\partial}{\partial x}$ folgen aus (2.2) die Gleichungen

$$\frac{d\Phi}{dt} + \lambda_V \Phi^2 + \lambda_W W_x \Phi = Y_u w + (Y_V - \lambda_u w)\Phi + Y_W W_x$$

$$\frac{dW}{dt} = Z + (\lambda - \mu) W_x .$$

Es seien F_1 und F_2 Funktionen von u, V und W, welche in $\mathfrak{R}$ zweimal bzw. einmal stetig differenzierbar sind und den Dgln.

$$\frac{\partial F_1}{\partial W} = \frac{\lambda_W}{\lambda - \mu}, \quad \frac{\partial F_2}{\partial W} = \frac{Y_W e^{F_1}}{\lambda - \mu}$$

genügen. Setzt man dann

$$(2.3)\qquad \Psi = e^{F_1}\Phi - F_2 ,$$

so erhält man eine Dgl.

$$(2.4)\qquad \frac{d\Psi}{dt} = G_0 + G_1\Psi + G_2\Psi^2 ,$$

worin G_0, G_1 und G_2 in $\mathfrak{R}$ stetige Funktionen sind mit

$$G_2 = -\lambda_V e^{-F_1} > 0 .$$

2) Mit der Abkürzung $\mathfrak{v} = (u, V, W)$ sei $\mathfrak{v}_0$ ein innerer Punkt von $\mathfrak{R}$ und $\mathfrak{R}_\varrho$: $|\mathfrak{v} - \mathfrak{v}_0| \leq \varrho$ eine Kugel derart, daß auch die Kugel $\mathfrak{R}_{2\varrho}$ noch zum Innern von $\mathfrak{R}$ gehört. In $\mathfrak{R}_{2\varrho}$ gelte

$$(2.5)\qquad \begin{aligned} &\Theta(\mathfrak{v}) = \{|X|^2 + |Y|^2 + |Z|^2\}^{1/2} \leq c_1 ; \\ &G_2 \geq c_2 > 0 ; \quad G_0, G_1, F_1 \geq c_3 ; \quad F_2 \leq c_4 . \end{aligned}$$

Eine in dem Bestimmtheitsgebiet

$$(2.6)\qquad \mathfrak{B}: |x| \leq 2\tau - t , \quad 0 \leq t < \tau , \quad \tau = \frac{\varrho}{c_1}$$

zweimal stetig differenzierbare Lösung $\mathfrak{v}(x, t)$ des Systems (2.2) mit Anfangswerten $\mathfrak{v}(x, 0) \in \mathfrak{R}_\varrho$ für $|x| \leq 2\tau$ *liegt ganz in* $\mathfrak{R}_{2\varrho}$. Dies folgt aus

$$|\mathfrak{v}(x, t) - \mathfrak{v}(x, 0)| \leq t_0 \max_{t \leq t_0} \Theta(\mathfrak{v}(x, t))$$

und aus (2.5), (2.6) mit Hilfe einer bekannten indirekten Schlußweise. Es gelten also die a-priori-Abschätzungen (2.5) für die Funktionen $G_i(\mathfrak{v}(x, t))$ und $F_i(\mathfrak{v}(x, t))$. Aus der Dgl. (2.4) folgt damit durch Integration längs der vom

13*

Punkt $x = 0$, $t = 0$ ausgehenden Charakteristik mit der Steigung λ, daß $\Psi(x, t)$ auf der Charakteristik größer ist als die Lösung $\omega(t)$ der Dgl.

$$\frac{d\omega}{dt} = c_3 + c_3\omega + c_2\omega^2$$

mit $\omega(0) = \Psi(0, 0)$. Diese strebt gegen $+\infty$ für $t \to t' < \tau$, falls $\omega(0)$ größer ist als eine von c_2, c_3 und τ abhängige Zahl c_5. Für eine in $\mathfrak{B}$ zweimal stetig differenzierbare Lösung $\mathfrak{v}(x, t)$ gilt also nach (2.3) und (2.5)

$$V_x(0, 0) = \Phi(0, 0) = e^{-F_1}(\Psi(0, 0) + F_2) \leqq e^{-c_1}(c_5 + c_4) = c\,.$$

3) Man wähle nun unendlich oft differenzierbare Anfangswerte u, v und w so aus, daß die transformierten Anfangswerte $\mathfrak{v}(x, 0)$ für $|x| \leqq 2\tau$ in $\mathfrak{K}_\varrho$ liegen und $V_x(0, 0) > c$ ist. Dann ist nach Teil 2) des Beweises die Lösung $\mathfrak{v}(x, t)$ in $\mathfrak{B}$ nicht überall zweimal stetig differenzierbar. Sei $0 \leqq t < t_0$ der breiteste Streifen von $\mathfrak{B}$, in dem $\mathfrak{v}$ noch zweimal stetig differenzierbar ist. Wenn die ersten Ableitungen $\mathfrak{v}_x$ in diesem Streifen gleichmäßig beschränkt wären, so könnte man nach einem bekannten Existenzsatz[5]) einen breiteren Streifen finden, in dem $\mathfrak{v}$ zweimal stetig differenzierbar ist. Also sind die zweiten Ableitungen der Lösung $u(x, t)$ der Wellengleichung in dem Streifen $0 \leqq t < t_0$ nicht alle beschränkt. Damit ist der Beweis vollständig.

Die im Verlauf des Beweises wesentlich benutzte Bedingung $\lambda\lambda_v - \lambda_w \neq 0$ besagt, nach einer Definition von P. D. LAX [9], daß die Wellengleichung in bezug auf die λ-Charakteristiken „eigentlich nichtlinear" (genuinely nonlinear) ist. Umgekehrt bedeutet $\lambda\lambda_v - \lambda_w \equiv 0$, daß die Wellengl. in bezug auf die λ-Charakteristiken „linear ausgeartet" (linearly degenerate) ist. Das Resultat dieses Paragraphen kann daher auch so formuliert werden: Wenn die Wellengl. nicht vollständig, d. h. in bezug auf beide Charakteristikenscharen, linear ausgeartet ist, so ist das Anfangswertproblem im großen im eigentlichen Sinne nicht lösbar. Die Wellengleichungen der ausgezeichneten Klasse, und nur diese, sind vollständig linear ausgeartet.

§ 3. Schwache Lösungen

Der Satz 2 hat gezeigt, daß das Anfangswertproblem im großen für die nicht ausgezeichneten Wellengleichungen im eigentlichen Sinne nicht lösbar ist. Es gibt stets Anfangswerte derart, daß die zweiten Ableitungen der Lösung bei Annäherung an einen im endlichen gelegenen Punkt unbeschränkt wachsen. Eine ähnliche Situation ist aus der Gasdynamik bekannt; dort werden die ersten Ableitungen der abhängigen Veränderlichen unendlich groß. Dennoch gibt es im Fall der Gasdynamik eine physikalisch richtige Lösung des Anfangswertproblems, die aber nur stückweise stetig und stetig differenzierbar ist und an den Unstetigkeitsstellen gewissen Sprungbedingungen genügt. Man bezeichnet sie als *schwache* Lösung; physikalisch ist sie als *Stoßwelle* zu interpretieren[6]). Überträgt man diesen Begriff sinngemäß auf die Wellengleichungen, so kommt man zu folgender

[5]) Vgl. P. D. LAX [7].

[6]) Vgl. z. B. R. COURANT, K. O. FRIEDRICHS [4].

Definition. ***Eine für $t \geqq 0$ stetige und stückweise zweimal stetig differenzierbare Funktion $u(x)$ heißt eine schwache Lösung der Wellengleichung, wenn sie in jedem Stetigkeitsgebiet der ersten Ableitungen der Wellengleichung genügt und auf jeder Unstetigkeitsfläche $\mathfrak{S}$ der ersten Ableitungen den Sprungbedingungen***

$$[u_i] = \sigma n_i \tag{3.1}$$

[7]

$$[L_K u^i] n_i = 0 \,. \tag{3.2}$$

Hierin ist n_i die durch $\sum_{i=0}^{3} (n_i)^2 = 1$ normierte Normale der Fläche $\mathfrak{S}$; $[u_i]$ bezeichnet die Differenz $u_i^+ - u_i^-$ der Grenzwerte der ersten Ableitungen auf den beiden Seiten der Fläche, wobei die Normale etwa zur $+$-Seite hinweisen soll; σ ist eine durch (3.1) definierte, stetige Funktion auf $\mathfrak{S}$.

Die Sprungbedingungen (3.1) sind eine Folge der Stetigkeit der Funktion $u(x)$. Die Sprungbed. (3.2) garantiert, daß $u(x)$ der Integralgleichung

$$\int_{t>0} \{L_K u^i \varphi_i + L_I u \varphi\} \, dx = 0 \tag{3.3}$$

genügt, wenn $\varphi(x)$ eine beliebige Funktion aus der folgenden Menge $\mathfrak{D}^1$ ist: $\varphi(x)$ ist stetig differenzierbar für alle x; $\varphi(x) \equiv 0$ außerhalb eines beschränkten Teilgebietes $\mathfrak{G}_\varphi$ des Halbraumes $t > 0$. Umgekehrt ist jede stetige und stückweise zweimal stetig differenzierbare Funktion $u(x)$, die für beliebiges $\varphi \in \mathfrak{D}^1$ der Integralgleichung (3.3) genügt, eine schwache Lösung der Wellengleichung.

Für eine schwache Lösung gelten die Erhaltungssätze der Energie und des Impulses (1.4) in jedem Stetigkeitsgebiet der ersten Ableitungen. Für die Gültigkeit der Erhaltungssätze in der Integralform (1.4a) sind jedoch die *zusätzlichen* Sprungbed.

$$[T^{ij}] n_j = 0 \tag{3.4}$$

notwendig und hinreichend. Diesen sind wiederum die Integralgln.

$$\int_{t>0} T^{ij} \varphi_j \, dx = 0 \,, \tag{3.5}$$

gültig für beliebige Funktionen $\varphi \in \mathfrak{D}^1$, äquivalent[8]).

Zur Motivierung der Gln. (3.3) und (3.5) sei noch bemerkt, daß diese Gln. auch aus dem Variationsprinzip abgeleitet werden können als notwendige Bedingungen dafür, daß das Funktional

$$J\{u\} = \int_{t>0} L(K\{u\}, u^2) \, dx \,,$$

dessen Eulersche Gleichung die Wellengleichung ist, für eine stetige und stückweise stetig differenzierbare Funktion $u(x)$ einen stationären Wert annimmt.

---

[7]) Mit den Bezeichnungen von § 1.

[8]) Benutzt man die Gl. (3.3) dazu, allgemeinere schwache Lösungen zu definieren, z. B. solche, die stetig und stückweise stetig differenzierbar sind, so sind die Gln. (3.4) zwar notwendig für die Gültigkeit der Erhaltungssätze in der Form (1.4a) oder (3.5); es ist aber nicht klar (und vielleicht auch nicht richtig), daß sie hinreichen.

Man erhält nämlich Gl. (3.3) mit Hilfe der Vergleichsfunktionen $u_\varepsilon(x) = u(x) + \varepsilon\varphi(x)$, $\varphi(x) \in \mathfrak{D}^1$. Die i-te Gl. (3.5) folgt bei Benutzung der Schar

$$u_\varepsilon(x) = u(y_\varepsilon(x)), \quad y_\varepsilon^j(x) = \begin{cases} x^i + \varepsilon\varphi(x) & \text{für } j = i, \\ x^j & \text{für } j \neq i. \end{cases}$$

Über die Existenz schwacher Lösungen wird im folgenden nichts ausgesagt. Es wird vielmehr angenommen, daß es eine Menge schwacher Lösungen mit gewissen Eigenschaften gibt; daraus wird dann eine notwendige und hinreichende Bedingung für die Gültigkeit der Erhaltungssätze hergeleitet.

Satz 3.1. *Zu beliebigen Zahlenpaaren* (K_+, I) *und* (K_-, I) *aus dem Definitionsbereich der Lagrangefunktion* L *gebe es eine schwache Lösung* $u(x)$ *und eine Unstetigkeitsstelle der ersten Ableitungen von* $u(x)$ *derart, daß dort* $u^2 = I$, $(u^i u_i)_+ = K_+$, *und* $(u^i u_i)_- = K_-$ *ist. Behauptung: Die Erhaltungssätze in der Integralform* (1.4a) *sind für alle diese Lösungen dann und nur dann erfüllt, wenn die Wellengleichung zur ausgezeichneten Klasse gehört.*

Beweis. Nach Voraussetzung gelten für die Werte u, u_i^+, u_i^- der Lösung und ihrer ersten Ableitungen an der Unstetigkeitsstelle und für die Normale n_i der Unstetigkeitsfläche $\mathfrak{S}$ in diesem Punkt die Sprungbed. (3.1) und (3.2) mit einer Zahl $\sigma \neq 0$. Für das Bestehen der Erhaltungssätze ist die Gültigkeit der Sprungbedingungen (3.4) notwendig und hinreichend. Nach (1.3) gilt

$$[T^{ij}]\, n_j = 2\,[L_K u^j n_j u^i] - [L]\, n^i\,.$$

Wegen (3.2) ist aber

$$L_K u^j n_j = \{\alpha L_K^+ u_+^j + \beta L_K^- u_-^j\} n_j\,,$$

wenn α, β zwei Zahlen mit $\alpha + \beta = 1$ sind. Wählt man insbesondere

$$\alpha = \frac{L_K^-}{L_K^+ + L_K^-}\,, \quad \beta = \frac{L_K^+}{L_K^+ + L_K^-}\,,$$

so folgt mit Benutzung von (3.1)

$$\begin{aligned}[T^{ij}]n_j &= 2\,\frac{L_K^+ L_K^-}{L_K^+ + L_K^-}\{u_+^j + u_-^j\} n_j [u^i] - [L] n^i \\ &= \left\{2\,\frac{L_K^+ L_K^-}{L_K^+ + L_K^-}(K_+ - K_-) - L^+ + L^-\right\} n^i\,.\end{aligned}$$

Setzt man nun

$$\Phi(K_+, K_-, I) = 2\,\frac{L_K^+ L_K^-}{L_K^+ + L_K^-}(K_+ - K_-) - L^+ + L^-\,,$$

so sind die vier Gln. (3.4) der einen Gl. $\Phi(K_+, K_-, I) = 0$ äquivalent. Nun kann man zeigen, daß die Funktion Φ dann und nur dann für alle Zahlenpaare (K_+, I) und (K_-, I) aus dem Definitionsbereich der Lagrangefunktion L verschwindet, wenn die Wellengleichung zur ausgezeichneten Klasse gehört. Man berechnet

$$\frac{\partial\Phi}{\partial K_+}(K_+, K_-, I) = \frac{(L_K^+)^2 (L_K^-)^2}{(L_K^+ + L_K^-)^2}\{M^+ - M^- - (K_+ - K_-)\, M_K^+\}$$

und

$$\frac{\partial^2\Phi}{(\partial K_+)^2}(K, K, I) = -\frac{1}{4}(L_K)^3 M_{KK}$$

mit $M = (L_K)^{-2}$. Ist $\Phi(K_+, K_-, I) = 0$, so folgt wegen $L_K \neq 0$, daß $M_{KK}(K, I)$ identisch verschwindet. Umgekehrt folgt aus $M_{KK} = 0$ zunächst $\frac{\partial \Phi}{\partial K_+} = 0$ und daraus $\Phi(K_+, K_-, I) = 0$ wegen $\Phi(K, K, I) = 0$. Da die ausgezeichneten Wellengleichungen durch $M_{KK} = 0$ charakterisiert sind, ist der Beweis hiermit beendet.

Eine weitere Charakterisierung der ausgezeichneten Wellengleichungen ergibt sich durch Vergleich der Unstetigkeitsflächen $\mathfrak{S}$ mit den Charakteristiken:

Satz 3.2. *Unter den Voraussetzungen von Satz* 3.1 *ist jede Unstetigkeitsfläche der betrachteten schwachen Lösungen dann und nur dann charakteristisch, wenn die Wellengleichung ausgezeichnet ist.*

Beweis. Die Unstetigkeitsfläche ist charakteristisch, wenn ihre Normale der Gleichung $A^{ij} n_i n_j = 0$ genügt, wo A^{ij} die Matrix der Koeffizienten des Hauptteils der Wellengleichung ist. Nach (1.7) gilt dann

$$n^i n_i = \frac{M_K}{M} (u^i n_i)^2 = M_K (L_K u^i n_i)^2$$

auf beiden Seiten der Fläche, also wegen (3.2):

$$[M_K (L_K u^i n_i)^2] = (L_K u^i n_i)^2 [M_K] = 0 .$$

Wäre nun $L_K u^i n_i = 0$ (auf einer Seite, also auch auf beiden), so folgte

$$\sigma(u^i_+ + u^i_-) n_i = K_+ - K_- = 0 .$$

Für $K_+ - K_- \neq 0$ ist also $[M_K] = M_K^+ - M_K^- = 0$, folglich M_K unabhängig von K, also $M_{KK} = 0$. Ist umgekehrt $M_{KK} = 0$, so gibt es nach § 1 zwei Möglichkeiten[9]):

1) Die Wellengleichung ist semilinear; in diesem Fall ist $n^i n_i = 0$ die Bedingung für die Normale einer charakteristischen Fläche. Andererseits ist $L_K = \text{const}$ und daher nach (3.1) und (3.2) $[u^i] n_i = \sigma n^i n_i = 0$ für eine Unstetigkeitsfläche $\mathfrak{S}$, d. h. $\mathfrak{S}$ ist beiderseits charakteristisch.

2) Die Wellengleichung ist vom Bornschen Typ. Setzt man hier

$$W = \sqrt{K + F(I)} ,$$

so ist

$$L_K = \frac{1}{2} W^{-1} , \quad M = 4 W^2 , \quad M_K = 4 .$$

Also ist nach (3.2)

$$\left[\frac{u^i n_i}{W}\right] = \frac{u^i_+ n_i}{W_+} - \frac{u^i_- n_i}{W_-} = 0$$

und andererseits nach (3.1)

$$\sigma u^i_+ n_i + \sigma u^i_- n_i = [K] = [W^2] = (W_+ + W_-) [W] .$$

---

[9]) Hier wird von der Tatsache Gebrauch gemacht, daß die in § 1 zur Rückführung der ausgezeichneten Wellengln. auf die semilinearen bzw. Bornschen Gln. benutzte Transformation schwache Lösungen der ursprünglichen Gl. in schwache Lösungen der transformierten Gl. überführt.

Daraus folgt

(3.6) $$\sigma u^i n_i = W[W]$$

auf beiden Seiten der Fläche und weiter

(3.6a) $$\sigma^2 n^i n_i = [W]^2 .$$

Die Gleichung der Charakteristiken

$$n^i n_i - \frac{M_K}{M}(u^i n_i)^2 = n^i n_i - \left(\frac{u^i n_i}{W}\right)^2 = 0$$

ist also auf beiden Seiten der Fläche erfüllt, q. e. d.

Zum Schluß des Paragraphen soll durch einen weiteren Satz die Verwandtschaft der ausgezeichneten Wellengln. mit den *linearen* hyperbolischen Dgln. noch deutlicher sichtbar gemacht werden.

Satz 3.3. *Bei den ausgezeichneten Wellengleichungen genügt die „Sprunghöhe" σ einer gewöhnlichen Differentialgleichung $\frac{d\sigma}{ds} + \alpha\sigma = 0$ längs jeder Bicharakteristik in einer Sprungfläche $\mathfrak{S}$ (s ist die Bogenlänge auf der Bicharakteristik, α eine stetige Funktion auf $\mathfrak{S}$). Die ersten Ableitungen einer schwachen Lösung sind demnach dann und nur dann unstetig, wenn ihre Anfangswerte für $t = 0$ es sind.*

Beweis. Für die semilinearen Gln. ist diese Tatsache bekannt[10]). Eine Bornsche Wellengleichung ist von der Form

$$a^{ij}u_{ij} + \{2W^2G' + W^{-1}FF' - 2WF'\}u = 0 ,$$

wobei $W = \sqrt{K + F(I)}$ gesetzt ist und

$$a^{ij} = Wg^{ij} - W^{-1}u^i u^j .$$

Eine Unstetigkeitsfläche ist nach Satz 3.2 charakteristisch, d. h. für ihre Normale gilt $a^{ij}n_i n_j = 0$. Die Richtung der Bicharakteristiken ist durch den Vektor

$$a^i = a^{ij}n_j = Wn^i - \sigma^{-1}[W]u_i$$

definiert, der wegen $a^i n_i = 0$ in der Fläche liegt. Aus (3.1) folgt $[a^i] = 0$, d. h. die Richtung der Bicharakteristiken auf den beiden Seiten der Fläche ist dieselbe. Nun berechnet man den Sprung der linken Seite der Wellengleichung:

$$[a^{ij}u_{ij}] + \{2(W_+ + W_-)G' - (W_+W_-)^{-1}FF' - 2F'\}[W]u = 0 .$$

Hierin läßt sich der zweite Term wegen (3.6a) als Produkt einer auf $\mathfrak{S}$ stetigen Funktion mit σ schreiben. Zur Ausrechnung des ersten Terms führt man neue Koordinaten ξ^i ein derart, daß die Fläche $\mathfrak{S}$ durch $\xi^0 = 0$ dargestellt wird. Sei $\tilde{u}(\xi) = u(x(\xi))$, $\gamma^i_j = \frac{\partial\xi^i}{\partial x^j}$; dann gilt $n_i = \gamma^0_i$, $[\tilde{u}_j] = \sigma\delta^0_j$, $[\tilde{u}_{ij}] = 0$ für $i, j \neq 0$

---

[10]) Vgl. R. COURANT, D. HILBERT [5], Kap. V.

und

$$[a^{ij}u_{ij}] = [a^{ij}(\gamma_i^k \gamma_j^m \tilde{u}_{km} + \gamma_{ij}^k \tilde{u}_k)]$$

$$= 2a^i \gamma_i^k [\tilde{u}_{ok}] + [a^{ij}]\gamma_i^k \gamma_j^m \tilde{u}_{km} + \frac{1}{2}[a^{ij}]\gamma_{ij}^k(\tilde{u}_k^+ + u_k^-) +$$

$$+ \frac{1}{2}(a_+^{ij} + a_-^{ij})\gamma_{ij}^k[\tilde{u}_k] = 2a^i\sigma_i + \beta\sigma ,$$

worin β eine auf $\mathfrak{S}$ stetige Funktion bedeutet. Dabei ist zu bemerken, daß in dem Ausdruck $[a^{ij}]\gamma_i^k\gamma_j^m\tilde{u}_{km}$ keine Differentiationen nach ξ^0 vorkommen, und daß $[a^{ij}] = \sigma b^{ij}$ gesetzt werden kann, wobei die b^{ij} auf $\mathfrak{S}$ stetig sind. Insgesamt erhält man also eine Dgl. der Form $a^i\sigma_i + \alpha\sigma = 0$, und es bleibt zu zeigen, daß man noch mit dem Ausdruck $\left\{\sum_{i=0}^{3}(a^i)^2\right\}^{1/2}$ dividieren kann. Nach (3.6) und (3.6a) ist $a^i a_i = -\sigma^{-2}[W]^2(W^2 - K) = -n^i n_i F(I)$. Ist $n^i n_i > 0$, so ist $a^i a_i < 0$ wegen $F > 0$, also gilt $\sum_{i=0}^{3}(a^i)^2 > 0$. Ist $n^i n_i = 0$, so hat man $a^i = W n^i$, $\sum_{i=0}^{3}(a^i)^2 = W^2 > 0$, q. e. d.

§ 4. Komplexwertige Wellenfunktionen

Die bisherigen Betrachtungen kann man, mit einer Ausnahme, auf Wellengleichungen für eine komplexwertige Funktion übertragen. Dazu geht man wieder von einer Lagrangefunktion $L = L(K, I)$ aus, jedoch mit

$$K = g^{ij}u_i\overline{u_j} = |u_x|^2 + |u_y|^2 + |u_z|^2 - |u_t|^2$$
$$I = u\overline{u} = |u|^2$$

und erhält dieselbe Form (1.2) der Wellengleichung. Führt man die Differentiation aus, so ergibt sich, im Unterschied zu (1.2a), der Hauptteil

$$\{L_K g^{ij} + L_{KK}u^i\overline{u^j}\}u_{ij} + L_{KK}u^i u^j \overline{u_{ij}} .$$

Die Normale einer charakteristischen Fläche muß demnach der Gleichung

$$L_K n^i n_i \{L_K n^j n_j + 2L_{KK}|u^i n_i|^2\} = 0 .$$

genügen. Die durch einen Punkt gehende charakteristische Mannigfaltigkeit zerfällt also in den Lichtkegel und ein (im abgeschlossenen Lichtkegel enthaltenes) Konoid. Die notwendigen und hinreichenden Bedingungen für den hyperbolischen bzw. für den kausalen Charakter der Wellengleichung, nämlich die Gln. (1.5) bzw. (1.6), bleiben unverändert. Die Definition (1.3) des Energie-Impuls-Tensors ist durch

$$T^{ij} = L_K(u^i\overline{u^j} + \overline{u^i}u^j) - Lg^{ij} \tag{4.1}$$

zu ersetzen. Aus der Invarianz der Lagrangefunktion gegenüber der Transformation $u \to e^{i\alpha}u$ (α reell) folgt der Erhaltungssatz

$$\frac{\partial}{\partial x^j}J^j \equiv \frac{\partial}{\partial x^j}\left\{\frac{1}{i}L_K(u\overline{u^j} - \overline{u}u^j)\right\} = 0 , \tag{4.2}$$

der als Erhaltungssatz der elektrischen Ladung zu interpretieren ist; J^0 ist die Ladungsdichte. Diesem Erhaltungssatz entspricht bei den schwachen Lösungen die Sprungbedingung

$$[J^j]\, n_j \equiv \frac{1}{i}\{u\,\overline{[L_K u^j]}\, n_j - \bar{u}\,[L_K u^j]\, n_j\} = 0\,,$$

die wegen (3.2) erfüllt ist.

Die Sätze 2 und 3.1 bleiben unverändert gültig. Für Satz 2 ist das offensichtlich; Satz 3.1 muß wegen der veränderten Definition des Energie-Impuls-Tensors neu verifiziert werden. Zur Übertragung von Satz 3.2 hat man zwischen dem Lichtkegel und dem charakteristischen Konoid zu unterscheiden. Man definiert: Eine Fläche heißt „charakteristisch im engeren Sinne", wenn sie in jedem Punkt das charakteristische Konoid berührt, d. h. wenn ihre Normale der Gleichung

$$n^i n_i - \frac{M_K}{M}\,|u^i n_i|^2 = 0$$

genügt. Ersetzt man nun in Satz 3.2 das Wort „charakteristisch" durch „charakteristisch im engeren Sinne", so bleibt der Satz auch für komplexwertige Wellenfunktionen richtig[11]). Der Beweis von Satz 3.3 läßt sich nur bei den semilinearen Wellengleichungen auf den Fall komplexwertiger Funktionen übertragen. Bei den Bornschen Wellengln. einer komplexwertigen Funktion existiert keine entsprechende Differentialgleichung für die Sprunghöhe σ.

§ 5. Das Anfangswertproblem für die semilinearen Wellengleichungen in zwei Variablen

In diesem und dem folgenden Paragraphen wird die bisherige Betrachtungsweise umgekehrt und versucht, das Anfangswertproblem im großen für die ausgezeichneten Wellengleichungen zu lösen. Es wird dazu die vereinfachende Annahme gemacht, daß die Anfangswerte nur von einer Variablen x abhängen, so daß die Lösung eine Funktion der zwei Variablen x, t ist. Die Lagrangefunktion einer semilinearen Wellengleichung für eine komplexwertige Funktion $u(x, t)$ ist nach § 1 und § 4 von der Form

$$L = \frac{1}{2}\{|u_x|^2 - |u_t|^2 + F(|u|^2)\}$$

und die Wellengleichung ist

$$u_{tt} - u_{xx} + F'(|u|^2)\,u = 0\,. \tag{5.1}$$

Die Funktion $F(I)$ sei für $0 \leq I < \infty$ zweimal stetig differenzierbar und so

[11]) Das bedeutet nicht, daß eine Charakteristik mit $n^i n_i = 0$ nicht Träger einer Unstetigkeit der ersten Ableitungen sein kann; eine Fläche kann nämlich zugleich Charakteristik der einen und der anderen Art sein. Ein Beispiel ist eine ins Vakuum (Gebiet mit $u \equiv 0$) hineinlaufende Stoßwelle.

beschaffen, daß das Energie-Integral

$$E_t\{u\} = \frac{1}{2}\int_{-\infty}^{+\infty} \{|u_x|^2 + |u_t|^2 + F(|u|^2)\}\,dx \tag{5.2}$$

für $u \not\equiv 0$ positiv ist, falls es existiert, und nur für $u \equiv 0$ verschwindet. Dazu ist notwendig und hinreichend, daß gilt

$$F(0) = 0\,,\quad F(I) > 0 \quad \text{für} \quad 0 < I < \infty\,.$$

Darüber hinaus soll noch die Ungleichung

$$\liminf_{I\to\infty} F(I) > 0$$

vorausgesetzt werden. Diese Ungleichung und der Energiesatz liefern eine a priori-Abschätzung für die Lösungen der Wellengleichung (5.1), aus der der Existenzsatz folgt.

Hilfssatz 5.1. *Es sei $u(x,t)$ eine zweimal stetig differenzierbare Lösung der Wellengleichung (5.1) für $t \geqq 0$ mit den Anfangswerten $u(x,0) = u_0(x)$, $u_t(x,0) = u_1(x)$, und es sei*

$$E_0\{u\} = \frac{1}{2}\int_{-\infty}^{+\infty} \{|u_0'(x)|^2 + |u_1(x)|^2 + F(|u_0(x)|^2)\}\,dx < \infty\,.$$

Dann existiert das Energie-Integral $E_t\{u\}$ für $t \geqq 0$, und es gilt $E_t\{u\} \leqq E_0\{u\}$, $|u(x,t)| \leqq C$ mit einer nur von $E_0\{u\}$ und der Funktion $F(I)$ abhängigen Zahl C.

Beweis. Der erste Teil der Behauptung ergibt sich aus dem Energiesatz

$$\{|u_x|^2 + |u_t|^2 + F(|u|^2)\}_t = \{u_x\overline{u}_t + \overline{u}_x u_t\}_x$$

durch Integration über das Gebiet $-\xi + t \leqq x \leqq \xi - t$, $0 \leqq t \leqq \tau$ und den Grenzübergang $\xi \to \infty$.

Es sei $\varrho = \liminf_{I\to\infty} F(I)$; nach Voraussetzung ist $\varrho > 0$. Ist $\mathfrak{M}$ die Menge aller x mit $F(|u(x,t)|^2) \leqq \frac{\varrho}{2}$, so ist $|u(x,t)| \leqq C_1$ für alle x aus $\mathfrak{M}$ und C_1 hängt nur von der Funktion $F(I)$ ab. $\mathfrak{M}$ ist nicht leer, denn das Maß der Komplementärmenge $\overset{1}{\mathfrak{M}}$ ist offenbar nicht größer als $4\varrho^{-1}E_0\{u\}$. Andererseits gilt für zwei Punkte x_1, x_2, die in einem Intervall aus $\overset{1}{\mathfrak{M}}$ liegen

$$|u(x_2,t) - u(x_1,t)| \leqq \left\{|x_2 - x_1| \int_{-\infty}^{+\infty} |u_x(x,t)|^2 dx\right\}^{1/2} \leqq \sqrt{\frac{8E_0^2\{u\}}{\varrho}}\,.$$

Also ist

$$|u(x,t)| \leqq C_1 + \sqrt{\frac{8}{\varrho}}\;E_0\{u\} \quad \text{q. e. d.}$$

Mit Hilfe der gewonnenen Abschätzung kann man nun die stetige Abhängigkeit der Lösung von den Anfangswerten zeigen.

Hilfssatz 5.2. *Seien $u(x, t)$, $v(x, t)$ zweimal stetig differenzierbare Lösungen der Wellengleichung für $t \geqq 0$ mit $E_0\{u\} \leqq E$, $E_0\{v\} \leqq E < \infty$. Für die Differenz $w(x, t) = u(x, t) - v(x, t)$ gilt dann*

$$|w(x, t)| \leqq (1 + Ct)\, e^{Ct} \left\{ \max_{|\xi - x| \leqq t} |w_0(\xi)| + \frac{1}{2} \int_{x-t}^{x+t} |w_1(\xi)|\, d\xi \right\}$$

mit $w_0(x) = w(x, 0)$, $w_1(x) = w_t(x, 0)$ und einer nur von E abhängigen Konstanten C.

Der Beweis beruht auf der für die Lösungen der Wellengl. gültigen Integralgleichung

$$\begin{aligned} u(x, t) &= \frac{1}{2}\{u_0(x + t) + u_0(x - t)\} + \frac{1}{2} \int_{x-t}^{x+t} u_1(\xi)\, d\xi - \\ &\quad - \frac{1}{2} \int_0^t d\tau \int_{x-t+\tau}^{x+t-\tau} F'(|u(\xi, \tau)|^2)\, u(\xi, \tau)\, d\xi \end{aligned} \tag{5.3}$$

und kann übergangen werden. Der Hilfssatz zeigt, daß es zu gegebenen Anfangswerten höchstens eine zweimal stetig differenzierbare Lösung geben kann. Nun soll gezeigt werden, daß es genau eine Lösung gibt.

Satz 5.3[12]**.** *Es sei $u_0(x)$ zweimal, $u_1(x)$ einmal stetig differenzierbar, und es sei $E_0\{u\} < \infty$. Dann gibt es genau eine zweimal stetig differenzierbare Lösung der Wellengleichung für $t \geqq 0$ mit diesen Anfangswerten auf der Geraden $t = 0$.*

Beweis. Es genügt offenbar, zu zeigen, daß eine solche Lösung in einem Streifen $0 \leqq t \leqq T$ existiert, dessen Breite T nur von $E_0\{u\}$ abhängt. Ist dies gelungen, so erhält man, nach Hilfssatz 5.1 und wegen der bereits bewiesenen Einzigkeit, die Lösung in $t \geqq 0$ durch fortgesetzte Anwendung des Existenzsatzes.

Man gewinnt eine Lösung der Wellengleichung durch iterative Auflösung der Integralgleichung (5.3). Dazu setzt man $u^{(0)}(x, t) \equiv 0$ und

$$u^{(n)}(x, t) = \frac{1}{2}\{u_0(x + t) + u_0(x - t)\} + \frac{1}{2} \int_{x-t}^{x+t} u_1(\xi)\, d\xi -$$

$$- \frac{1}{2} \int_0^t d\tau \int_{x-t+\tau}^{x+t-\tau} F'(|u^{(n-1)}|^2)\, u^{(n-1)}(\xi, \tau)\, d\xi$$

für $n = 1, 2, 3, \ldots$.

Nach Hilfssatz 5.1 ist $|u_0(x)| \leqq C_2$; C_2 eine nur von $E_0\{u\}$ abhängige Zahl. Mit

$$C_3 = \max_{0 \leqq I \leqq 2C_2} |F'(I)|$$

---

[12]) Ein analoger Satz für die (semilineare) Mesonengleichung von Schiff ist früher bereits von J. Moser bewiesen worden (mündliche Mitteilung von Herrn Moser, nicht veröffentlicht).

und der Induktionsannahme $|u^{(n-1)}(x, t)| \leq 2C_2$ erhält man

$$|u^{(n)}(x, t)| \leq C_2 + \left\{\frac{t}{2} \int_{-\infty}^{+\infty} |u_1(\xi)|^2 d\xi\right\}^{1/2} + C_2 C_3 t^2 \leq 2C_2$$

für $0 \leq t \leq T$ mit $T = \min\left(\frac{C_2^2}{4E_0\{u\}}, \frac{1}{\sqrt{2C_3}}\right)$. Also gilt $|u^{(n)}(x, t)| \leq 2C_2$ für $n = 1, 2, \ldots$. Setzt man nun

$$C_4^2 = C_3 + 4C_2 \cdot \max_{0 \leq I \leq 2C_2} |F''(I)|,$$

so folgt weiter

$$|u^{(n+1)}(x, t) - u^{(n)}(x, t)| \leq 2C_2 \frac{(C_4 t)^{2n}}{(2n)!}.$$

Damit ist die gleichmäßige Konvergenz der Folge für $0 \leq t \leq T$ bewiesen. Die Grenzfunktion ist stetig und genügt der Integralgleichung; also ist sie auch zweimal stetig differenzierbar und löst die Wellengleichung. Da T nur von $E_0\{u\}$ abhängt, ist der Beweis erbracht.

Von einer physikalisch sinnvollen Lösung muß verlangt werden, daß die Integrale $E_t\{u\}$ und

$$P_t\{u\} = -\int_{-\infty}^{+\infty} \mathrm{Re}(u_x \overline{u_t})\, dx$$

$$Q_t\{u\} = -\int_{-\infty}^{+\infty} \mathrm{Im}(u \overline{u_t})\, dx$$

(Energie, Impuls und Ladung) für $t \geq 0$ existieren und von t unabhängige Werte haben. Dies ist nun nicht nur für die in Satz 5.3 konstruierte Lösung, sondern für viel allgemeinere Lösungen der Fall, wie der folgende Satz zeigt.

Satz 5.4. *Sei $u_0(x)$ totalstetig und $u_1(x)$ integrierbar; ferner gelte $E_0\{u\} < \infty$. Dann gibt es eine für $t \geq 0$ definierte Funktion $u(x, t)$ mit den Eigenschaften:*

1) *u ist stetig und Lösung der Integralgleichung* (5.3) *für $t \geq 0$.*

2) *Die Integrale $E_t\{u\}$, $P_t\{u\}$ und $Q_t\{u\}$ existieren und hängen von t nicht ab.*

3) *Für jede Funktion $\varphi \in \mathfrak{D}^1$* (vgl. § 3) *sind die Integralgleichungen* (3.3) *und* (3.5) *sowie die analoge, aus* (4.2) *folgende, Gleichung erfüllt.*

Der Beweis, der nur angedeutet wird, beruht auf der Approximation durch zweimal stetig differenzierbare Lösungen, deren Anfangswerte außerhalb eines endlichen Intervalls (a, b) identisch verschwinden. Solche Lösungen sind identisch Null für $x \leq a - t$ und für $x \geq b + t$, wie man aus dem Beweis von Satz 5.3 sieht. Die Integrale $E_t\{u\}$, $P_t\{u\}$ und $Q_t\{u\}$ existieren daher und sind von t unabhängig; dies folgt aus den Erhaltungssätzen (1.4) bzw. (4.2). Die Konvergenz des Prozesses folgt aus

Hilfssatz 5.5. *Bezeichnet $\|f\|$ die L_2-Norm in bezug auf $-\infty < x < \infty$, so gilt unter den Voraussetzungen von Hilfssatz* 5.2

$$\max_x |w(x, t)| \leq (1 + Ct)e^{Ct}\left\{\max_\xi |w_0(\xi)| + \sqrt{\frac{t}{2}}\, \|w_1\|\right\}$$

und für beliebige Intervalle (a, b)

$$\left.\begin{array}{l}\left[\int\limits_a^b |w_x(x,t)|^2\,dx\right]^{1/2} \\ \left[\int\limits_a^b |w_t(x,t)|^2\,dx\right]^{1/2}\end{array}\right\} \leqq \|w_0'\| + \|w_1\| + \\ \qquad + \sqrt{b-a}\,Ct(1+Ct)e^{Ct}\left\{\max_\xi |w_0(\xi)| + \sqrt{\frac{t}{2}}\,\|w_1\|\right\}$$

mit der Konstanten C *von Hilfssatz* 5.2.

§ 6. Das Anfangswertproblem für die Bornschen Wellengleichungen in zwei Variablen

Die Lagrangefunktion einer Bornschen Wellengl. für eine reelle Funktion $u(x, t)$ ist nach § 1 von der Form

$$L = \sqrt{u_x^2 - u_t^2 + F(u^2)} - G(u^2)$$

mit $F > 0$. Im folgenden wird nur für den Fall $G \equiv 1$ ein Existenzsatz bewiesen. Über die Funktion $F(I)$ wird dabei vorausgesetzt:

$$F(I), F'(I), F''(I) \quad \text{stetig für} \quad 0 \leqq I < \infty,$$
$$F(0) = 1, F(I) > 1 \quad \text{für} \quad 0 < I < \infty$$
$$\text{und} \quad \liminf_{I\to\infty} F(I) > 1\,.$$

Die Wellengleichung ist

$$\left(\frac{u_t}{W}\right)_t - \left(\frac{u_x}{W}\right)_x + \frac{F'(u^2)\,u}{W} = 0\,, \tag{6.1}$$

wobei $W = \sqrt{u_x^2 - u_t^2 + F(u^2)}$ gesetzt ist.

In dem Beweis des Existenzsatzes dieses Paragraphen spielen der Energiesatz und die gleichmäßige Beschränktheit der Lösung nur eine geringe Rolle; diese Dinge werden daher ohne Beweis vorangeschickt.

Hilfssatz 6.1. *Sei* $u(x, t)$ *eine zweimal stetig differenzierbare Lösung der Wellengleichung* (6.1) *für* $t \geqq 0$ *mit* $u_x^2 - u_t^2 + F(u^2) > 0$. *Die Integrale*

$$E_t\{u\} = \int\limits_{-\infty}^{+\infty} \left\{\frac{u_x^2 + F(u^2)}{W} - 1\right\} dx$$

$$P_t\{u\} = -\int\limits_{-\infty}^{+\infty} \frac{u_x u_t}{W}\,dx$$

(Energie und Impuls) *existieren und sind unabhängig von* t, *falls* $E_0\{u\} < \infty$ *ist. Ferner ist* $|u(x, t)| \leqq C$ *mit einer nur von* $E_0\{u\}$ *abhängigen Konstanten* C.

Der Beweis benutzt wesentlich die Voraussetzungen $F(I) > 1$ und $\liminf\limits_{I\to\infty} F(I) > 1$; er ist nicht viel komplizierter als der von Hilfssatz 5.1.

Die Lösung der Wellengl. (6.1) wird in charakteristischen Variablen durchgeführt. Dazu führt man zunächst die neue abhängige Variable

$$z = \int_0^u \frac{ds}{\sqrt{F(s^2)}} \tag{6.2}$$

ein und setzt

$$\Theta(z) = \sqrt{F(u^2(z))}\,, \quad \frac{\Theta'(z)}{\Theta(z)} = \Phi(z)\,,$$

ferner

$$\begin{aligned} p &= \operatorname{arc\,tg} \frac{z_t}{\sqrt{1+z_x^2-z_t^2}} + \operatorname{arc\,tg} z_x \\ q &= \operatorname{arc\,tg} \frac{z_t}{\sqrt{1+z_x^2-z_t^2}} - \operatorname{arc\,tg} z_x\,. \end{aligned} \tag{6.3}$$

Es gilt

$$W = \Theta(z)\sqrt{1+z_x^2-z_t^2} = \Theta(z)\frac{\cos\frac{p+q}{2}}{\cos\frac{p-q}{2}}\,.$$

Die obige Transformation ist genau dann eineindeutig und stetig, wenn $W > 0$ ist, d. h. wenn $\cos p + \cos q > 0$ ist. Die Bedeutung dieser Ungleichung ergibt sich daraus, daß $\lambda = \cos p$, $\mu = -\cos q$ die charakteristischen Geschwindigkeiten sind. Nun werden charakteristische Variable α, β eingeführt derart, daß α auf den μ-Charakteristiken, β auf den λ-Charakteristiken konstant ist[13]). Man erhält das folgende System von Differentialgleichungen

$$\begin{aligned} & x_\alpha - t_\alpha \cos p = 0 && \mathrm{I} \\ & x_\beta + t_\beta \cos q = 0 && \mathrm{II} \\ & z_\alpha - t_\alpha \sin p = 0 && \mathrm{III} \\ & z_\beta - t_\beta \sin q = 0 && \mathrm{IV} \\ & q_\alpha + \frac{1}{2}(\cos p + \cos q)\,\Phi(z) t_\alpha = 0 && \mathrm{V} \\ & p_\beta + \frac{1}{2}(\cos p + \cos q)\,\Phi(z) t_\beta = 0\,. && \mathrm{VI} \end{aligned} \tag{6.4}$$

Daraus folgt weiterhin die Gleichung

$$t_{\alpha\beta} + \frac{1}{2}(\sin p + \sin q)\,\Phi(z) t_\alpha t_\beta = 0\,, \qquad \mathrm{VII}$$

die für len Existenzbeweis entscheidend ist. Die Anfangswerte seien auf der Geraden $\alpha + \beta = 0$ als Funktionen von α gegeben; und zwar darf man ohne Beschränkung der Allgemeinheit

$$x(\alpha, -\alpha) = \alpha\,, \quad t(\alpha, -\alpha) = 0$$

[13]) x, t, z, p, und q werden als Funktionen von α und β aufgefaßt. Vgl. H. Lewy [10] und R. Courant, D. Hilbert [5], Kap. V.

setzen. Zwischen $z(\alpha) = z(\alpha, -\alpha)$ und $p(\alpha)$, $q(\alpha)$ besteht die Beziehung

$$z'(\alpha) = \operatorname{tg} \frac{p(\alpha) - q(\alpha)}{2} \tag{6.5}$$

wegen (6.3). Außerdem berechnet man leicht

$$t_\alpha(\alpha, -\alpha) = t_\beta(\alpha, -\alpha) = \frac{1}{\cos p(\alpha) + \cos q(\alpha)}\,. \tag{6.6}$$

Satz 6.2. *Die Funktionen $z(\alpha)$, $p(\alpha)$, $q(\alpha)$ seien für $-\infty < \alpha < \infty$ erklärt; $z(\alpha)$ sei zweimal stetig differenzierbar und gleichmäßig beschränkt, $p(\alpha)$ und $q(\alpha)$ seien stetig differenzierbar und es sei $\cos p(\alpha) + \cos q(\alpha) > 0$. Ferner gelte die Gl.* (6.5). *Dann gibt es ein an die Gerade $\alpha + \beta = 0$ anschließendes Teilgebiet $\mathfrak{G}$ der Halbebene $\alpha + \beta \geqq 0$ und dort erklärte Funktionen x, t, z, p, q mit den Eigenschaften:*

1) *$\mathfrak{G}$ ist ein Bestimmtheitsgebiet für das System* (6.4), *d. h. ist (α_0, β_0) ein innerer Punkt von $\mathfrak{G}$, so gehört auch das Dreieck $\alpha \leqq \alpha_0$, $\beta \leqq \beta_0$, $\alpha + \beta \geqq 0$ ganz zu $\mathfrak{G}$.*

2) *$x(\alpha, \beta)$, $t(\alpha, \beta)$ und $z(\alpha, \beta)$ sind in $\mathfrak{G}$ zweimal, $p(\alpha, \beta)$ und $q(\alpha, \beta)$ einmal stetig differenzierbar und diese Funktionen stellen die einzige Lösung des Systems* (6.4) *mit den vorgeschriebenen Anfangswerten dar.*

3) *Das Bild des Gebietes $\mathfrak{G}$ bei der Abbildung*

$$(\alpha, \beta) \to (x(\alpha, \beta), t(\alpha, \beta))$$

ist die Halbebene $t \geqq 0$.

Bemerkung: Die Voraussetzung, daß $z(\alpha)$ beschränkt sei, kann man nach Hilfssatz 6.1 weglassen, wenn man die Existenz des Energie-Integrals voraussetzt. Dieses ist in den neuen Variablen gleich

$$\int_{-\infty}^{+\infty} \left\{ \frac{2\Theta(z(\alpha))}{\cos p(\alpha) + \cos q(\alpha)} - 1 \right\} d\alpha\,.$$

Wegen $\liminf\limits_{|z| \to \infty} \Theta(z) > 1$ und mit Hilfe von (6.5) kann man die Beschränktheit von $z(\alpha)$ hieraus auch direkt leicht beweisen.

Beweis von Satz 6.2. 1) Zunächst zeigt man, daß in dem System (6.4) je eine der Gln. I, II und der Gln. III, IV weggelassen werden können. Läßt man etwa II und IV weg, so seien $A(\alpha, \beta)$, $B(\alpha, \beta)$ die linken Seiten von II und IV. Man erhält $A_\alpha = 0$, $B_\alpha = 0$ und $A(\alpha, -\alpha) = 0$, $B(\alpha, -\alpha) = 0$ aus den übrigen Gln. und aus den Anfangsbedingungen (6.5), (6.6). Also sind II und IV auch erfüllt.

2) Das verkleinerte System besteht aus fünf Gln. für die fünf unbekannten Funktionen und ist im kleinen in bekannter Weise lösbar, d. h. zu jedem Intervall $\alpha_1 \leqq \alpha \leqq \alpha_2$ gibt es eine Zahl $\gamma > 0$, so daß in dem Trapez $\alpha \leqq \alpha_2$, $\beta \leqq -\alpha_1$, $0 \leqq \alpha + \beta \leqq \gamma$ eine und nur eine Lösung mit den vorgeschriebenen

Anfangswerten existiert, welche die behaupteten Differenzierbarkeitseigenschaften hat. Indem man diesen Prozeß so weit als möglich fortsetzt, gelangt man zu einem Existenzgebiet $\mathfrak{G}$, das offenbar ein Bestimmtheitsgebiet ist. Es ist jetzt nur noch die Behauptung 3) zu beweisen.

3) Die Funktionen t_α und t_β sind im Innern von $\mathfrak{G}$ positiv, wie man aus VII und (6.6) abliest. $t(\alpha, \beta)$ wächst also monoton auf jeder Geraden in $\mathfrak{G}$ mit nichtnegativer Steigung, und zwar geht t gegen ∞ bei Annäherung an einen Randpunkt von $\mathfrak{G}$. Wäre dies nicht so, so gäbe es einen Randpunkt (α_0, β_0) von $\mathfrak{G}$ derart, daß $t(\alpha, \beta)$ in dem Dreieck $\alpha \leqq \alpha_0$, $\beta \leqq \beta_0$, $\alpha + \beta \geqq 0$ gleichmäßig beschränkt wäre.

4) Zunächst wird der Fall eines beschränkten Dreiecks behandelt. Das Dreieck heiße $\mathfrak{D}$. In $\mathfrak{D}$ gilt $0 \leqq t(\alpha, \beta) \leqq T$ und folglich nach Gl. III bzw. IV

$$|z(\alpha, \beta)| \leqq |z(-\beta)| + T \leqq C_1$$

wegen $t_\alpha > 0$ bzw. $t_\beta > 0$. Also gilt auch $|\Phi(z(\alpha, \beta))| \leqq C_2$ und es folgt aus VII

$$0 \leqq t_\alpha(\alpha, \beta) \leqq t_\alpha(\alpha, -\alpha)\, e^{C_2 T} \leqq C_3 e^{C_2 T}$$
$$0 \leqq t_\beta(\alpha, \beta) \leqq C_3 e^{C_2 T},$$

wenn C_3 eine Schranke für den Ausdruck $\dfrac{1}{\cos p(\alpha) + \cos q(\alpha)}$ in dem Intervall $-\beta_0 \leqq \alpha \leqq \alpha_0$ bezeichnet. Damit folgt nun aus dem System die Stetigkeit der Funktionen x, t, z, p, q und der in Behauptung 2) des Satzes genannten Ableitungen in dem abgeschlossenen Dreieck $\overline{\mathfrak{D}}$. Man erhält einen Widerspruch gegen die Annahme, wenn man nun noch zeigt, daß die Lösung über $\overline{\mathfrak{D}}$ hinaus fortsetzbar ist. Es genügt, die Fortsetzung über eine Seite hinaus, etwa die Seite $\alpha = \alpha_0$, durchzuführen. Ohne Beschr. der Allgemeinheit sei $\alpha_0 = 0$ und die Lösung sei auch in einem (kleinen) Dreieck $0 \leqq \alpha \leqq \alpha_1$, $\beta \leqq 0$, $\alpha + \beta \geqq 0$ schon bekannt. Die Lösung soll nun in einem Rechteck $\mathfrak{R}$: $0 \leqq \alpha \leqq \alpha_2 \leqq \alpha_1$, $0 \leqq \beta \leqq \beta_0$ konstruiert werden. Man setzt in $\mathfrak{R}$: $z_0 = v_0 = w_0 = q_0 = p_0 = 0$ und

$$z_{n+1}(\alpha, \beta) = z(0, \beta) + \int_0^\alpha v_n \sin p_n \, d\alpha'$$

$$v_{n+1}(\alpha, \beta) = t_\alpha(\alpha, 0) \exp\left\{-\frac{1}{2}\int_0^\beta (\sin p_n + \sin q_n)\, \Phi(z_n) w_n \, d\beta'\right\}$$

$$w_{n+1}(\alpha, \beta) = t_\beta(0, \beta) \exp\left\{-\frac{1}{2}\int_0^\alpha (\sin p_n + \sin q_n)\, \Phi(z_n)\, v_n \, d\alpha'\right\}$$

$$q_{n+1}(\alpha, \beta) = q(0, \beta) - \frac{1}{2}\int_0^\alpha (\cos p_n + \cos q_n)\, \Phi(z_n) v_n \, d\alpha'$$

$$p_{n+1}(\alpha, \beta) = p(\alpha, 0) - \frac{1}{2}\int_0^\beta (\cos p_n + \cos q_n)\, \Phi(z_n) w_n \, d\beta'$$

für $n = 0, 1, 2, \ldots$

Die Funktionen z_1, v_1, w_1 haben eine gemeinsame Schranke C_4 in $\mathfrak{R}$. Bezeichnet man nun mit $\| \ \|$ die Maximum-Norm in $\mathfrak{R}$, so gelten die Abschätzungen

$$\|z_{n+1}\| \leqq C_4 + \alpha_2 \|v_n\|$$
$$\|v_{n+1}\| \leqq C_4 \exp(\beta_0 \|\Phi(z_n) w_n\|)$$
$$\|w_{n+1}\| \leqq C_4 \exp(\alpha_2 \|\Phi(z_n) v_n\|) .$$

Sei nun

$$C_5 = \max\left\{\frac{1}{2C_4}, \max_{|z| \leqq 2C_4} |\Phi(z)|\right\}$$

$$\alpha_2 = \min\{\alpha_1, (2C_4C_5)^{-1} \exp(-2\beta_0 C_4 C_5)\} .$$

Aus den (für $n = 1$ erfüllten) Induktionsannahmen

$$\|z_n\| \leqq 2C_4, \|v_n\| \leqq C_4 \exp(2\beta_0 C_4 C_5), \|w_n\| \leqq 2C_4$$

folgen dann dieselben Abschätzungen für z_{n+1}, v_{n+1} und w_{n+1}. Also gelten die Ungleichungen für alle n. Mit Hilfe dieser Ungleichungen zeigt man nun leicht die gleichmäßige Konvergenz der Folgen z_n, v_n, w_n, p_n, q_n in $\mathfrak{R}$. Für die Grenzfunktionen gilt $w_\alpha(\alpha, \beta) = v_\beta(\alpha, \beta)$. Definiert man dann noch

$$t(\alpha, \beta) = t(0, \beta) + \int_0^\alpha v(\alpha', \beta)\, d\alpha'$$

und

$$x(\alpha, \beta) = x(0, \beta) + \int_0^\alpha v \cos p \, d\alpha' ,$$

so hat man eine Lösung des Systems (6.4) in dem größeren Dreieck $\alpha \leqq \alpha_2$, $\beta \leqq \beta_0$, $\alpha + \beta \geqq 0$ konstruiert.

5) Es bleibt der Fall, daß $t(\alpha, \beta)$ in einem unendlichen Bestimmtheitsdreieck gleichmäßig beschränkt ist. Sei etwa $t(\alpha, \beta) \leqq T$ in $\beta \leqq \beta_0$, $\alpha + \beta \geqq 0$. Dann folgt aus Gl. III, daß auch $|z(\alpha, \beta)| \leqq C + T$ ist, wo C eine Schranke für $|z(\alpha)|$ bedeutet. Also ist $|\Phi(z(\alpha, \beta))| \leqq C_6$ in dem Dreieck. Dann ist aber nach Gl. VII und wegen (6.6) die Funktion $t(\alpha, \beta)$ nicht kleiner als die Lösung $s(\alpha, \beta)$ der Dgl.

$$s_{\alpha\beta} + C_6 s_\alpha s_\beta = 0$$

mit den Anfangswerten

$$s(\alpha, -\alpha) = 0 , \quad s_\alpha(\alpha, -\alpha) = s_\beta(\alpha, -\alpha) = \frac{1}{\cos p(\alpha) + \cos q(\alpha)} .$$

Diese Funktion kann man berechnen; es ist

$$s(\alpha, \beta) = \frac{1}{C_6} \log\left\{1 + C_6 \int_{-\beta}^{\alpha} \frac{d\gamma}{\cos p(\gamma) + \cos q(\gamma)}\right\}$$

und folglich ist $t(\alpha, \beta)$ nicht beschränkt im Widerspruch zur Annahme.

6) Nun kann man beweisen, daß das Bild des Gebietes $\mathfrak{G}$ bei der Abbildung $(\alpha, \beta) \to (x(\alpha, \beta), t(\alpha, \beta))$ die Halbebene $t \geqq 0$ ist. Man wähle eine beliebige Zahl $T > 0$; dann gibt es auf jeder Geraden $\alpha = \text{const}$ und auf jeder Geraden $\beta = \text{const}$ in $\mathfrak{G}$ einen ersten Punkt mit $t(\alpha, \beta) = T$, wenn man die Gerade von der Anfangsgeraden aus durchläuft. Diese Punkte bilden eine Kurve $\mathfrak{C}$ in $\mathfrak{G}$. Auf den Geraden $\alpha = \alpha_0$ ist nach Gl. II

$$x(\alpha_0, \beta) \geqq \alpha_0 - t(\alpha_0, \beta) \geqq \alpha_0 - T$$

und auf den Geraden $\beta = \beta_0$ nach I

$$x(\alpha, \beta_0) \leqq -\beta_0 + t(\alpha, \beta_0) \leqq -\beta_0 + T\,.$$

Das Bild des durch $\mathfrak{C}$, durch die Anfangsgerade und die Geraden $\alpha = \alpha_0 > T$ bzw. $\beta = \beta_0 > T$ berandeten Teilgebietes von $\mathfrak{G}$ enthält also das Rechteck $0 \leqq t \leqq T$, $-\beta_0 + T \leqq x \leqq \alpha_0 - T$. Läßt man nun α_0, β_0 und T gegen ∞ streben, so erhält man das gewünschte Resultat. Damit ist Satz 6.2 vollständig bewiesen.

Die in Satz 6.2 konstruierte Lösung des Systems (6.4) definiert nach (6.2) eine unter Umständen mehrdeutige Funktion $u(x, t)$ für $t \geqq 0$; jedoch ist $u(x, t)$ überall dort zweimal stetig differenzierbar, wo

$$x_\alpha t_\beta - x_\beta t_\alpha = (\cos p + \cos q)\, t_\alpha t_\beta \neq 0\,,$$

also $\cos p + \cos q \neq 0$ ist. Dort ist $u(x, t)$ Lösung der Wellengl. (6.1). An den Stellen mit $\cos p + \cos q = 0$ ist nach (6.3) entweder $u_x^2 - u_t^2 + F = 0$ oder u_x ist unstetig. Ist überall $\cos p + \cos q > 0$, so ist $u(x, t)$ eine eindeutige Funktion und Lösung der Wellengleichung (6.1); es scheint nicht leicht zu sein, Bedingungen für die Anfangswerte $u_0(x)$, $u_1(x)$ anzugeben derart, daß dies der Fall ist. Im Sonderfall $F(I) \equiv 1$[14]) kann man Bedingungen angeben; hier ist $\Phi(z) \equiv 0$ also nach Gln. V und VI $p(\alpha, \beta) = p(\alpha)$, $q(\alpha, \beta) = q(-\beta)$, und folglich müssen die Anfangswerte $p(\alpha)$, $q(\alpha)$ die Bedingung

$$\cos p(\alpha) + \cos q(-\beta) > 0$$

für alle Wertepaare α, β mit $\alpha + \beta \geqq 0$ erfüllen.

Literatur

[1] Blochinzew, D. I.: Die Fortpflanzung von Signalen in der nichtlinearen Feldtheorie. Dokl. Akad. Nauk SSSR. **82**, 553—556 (1952). — [2] Blochinzew, D. I., u. W. Orlow: J. exp. Theor. Physik **25**, 513 (1953). — [3] Born, M.: On the Quantum Theory of the Electromagnetic Field. Proc. roy. Soc. London, A **143**, 410—437 (1934). — [4] Courant, R., and K. O. Friedrichs: Supersonic Flow and Shock Waves. New York 1948. — [5] Courant, R., u. D. Hilbert: Methoden der Mathematischen Physik, Bd. II.

[14]) In diesem Fall kann die Lösung explizit angegeben werden; diese Lösung ist auch von T. Taniuti [12] gefunden worden.

Berlin 1937. — [6] Heisenberg, W.: Mesonenerzeugung als Stoßwellenproblem. Z. Physik 133, 65—79 (1952). — [7] Lax, P. D.: Nonlinear hyperbolic equations. Comm. pure appl. Math. 6, 231—258 (1953). — [8] Lax, P. D.: Initial value problems for nonlinear hyperbolic equations. Conference on Partial Differential Equations, Univ. of Kansas, Summer 1954. — [9] Lax, P. D.: The initial value problem for nonlinear hyperbolic equations in two independent variables. Contributions to the Theory of Partial Differential Equations, Princeton 1954. — [10] Lewy, H.: Über das Anfangswertproblem einer hyperbolischen nichtlinearen partiellen Differentialgleichung zweiter Ordnung mit zwei unabhängigen Veränderlichen. Math. Ann. 98, 179—191 (1927). — [11] Schiff, L. I.: Nonlinear meson theory of nuclear forces, I. Physic. Rev. 84, 1—9 (1951). — [12] Taniuti, T.: On the wave propagation in the nonlinear fields. Prog. Theor. Physics 17, 461—486 (1957).

(Eingegangen am 6. März 1959)

HARARY, F.
Math. Annalen 138, 203—210 (1959)

The Number of Functional Digraphs*

By

FRANK HARARY in Princeton, N. J.

1. Introduction

Our object is to find a generating function whose coefficients give the number of isomorphism types of functional digraphs with a given finite number of points. It will be seen that functional digraphs correspond to functions which are fixed-point-free. DAVIS [1] has already found an explicit formula for the number of types of functions on a finite set, and his methods may be readily adapted to the solution of this variation of his problem. However, in the process of deriving this generating function, we find certain geometric properties of functional digraphs which are of independent interest. We shall see that a finite functional digraph is constructible from directed cycles and rooted trees. In addition, we obtain a geometric characterization of infinite functional digraphs which has applications in recursion theory; see MYHILL [6].

In addition to enumerating all functional digraphs, using the elegant enumeration methods of PÓLYA [7], we consider functional digraphs having the three different kinds of connectedness proposed in [2] and enumerate these. We conclude by deriving the generating function for isomorphism classes of functions.

2. Geometric properties of functional digraphs

A *directed graph* or *digraph* D consists of a collection of *points* $a, b, c, \ldots$ together with certain ordered pairs of distinct points. Each ordered pair (a, b) of points in the digraph is called a *directed line* or briefly a *line*, and is denoted $\overrightarrow{ab}$. We note that by definition a digraph doesn't contain any line of the form $\overrightarrow{bb}$ from a point to itself. We later want to consider digraphs in which such lines, called *slings*, are admissible. We shall always regard a digraph as finite except where we actually call it infinite. As a general reference for the theory of graphs and directed graphs, finite and infinite, see KÖNIG [5].

The *out-degree* of a point b of digraph D is the number of points c such that the line $\overrightarrow{bc}$ is in D; the in-degree is defined similarly. A digraph is *functional* if every point has out-degree 1. The concept of a functional digraph arises in a psychological context in the study of the structure of a group of people in which each member extends exactly one invitation to another member.

* This work was supported by a grant from the Office of Naval Research to the Logistics Project at Princeton University. I wish to thank J. MYHILL for discussions on Section 2.

A (*directed*) *path* from a_0 to a_n consists of a collection of distinct points a_0, a_1, $a_2, \ldots, a_n$ together with the lines $\overrightarrow{a_0a_1}, \overrightarrow{a_1a_2}, \ldots, \overrightarrow{a_{n-1}a_n}$. A (directed) *cycle* is obtained from a path from a_0 to a_n on adding the line $\overrightarrow{a_na_0}$. If D is an infinite digraph, then a *ray* from a_0 consists of distinct points $a_0, a_1, a_2, \ldots$ together with the lines $\overrightarrow{a_0a_1}, \overrightarrow{a_1a_2}, \ldots$.

The following definitions of three kinds of connectedness for a digraph were given in [2]. A digraph is *strongly connected* or *strong* if for any two points b and c, there is a path from b to c (and hence also one from c to b). A *unilateral* digraph has a path from b to c *or* a path from c to b. Finally, a digraph is *weak* if there is no partition of the set P of all points into two nonempty disjoint subsets P_1 and P_2 such that there is no line joining a point of P_1 with one of P_2. A digraph is *disconnected* if it is not even weak. A *weak component* of a digraph is a maximal weak subgraph.

If S is any set of points in D, then *the subgraph* $\langle S\rangle$ *generated by* S has S as its set of points and consists of those directed lines of D joining two points of S. A point c is *accessible* or *reachable* from b if there is a path in D from b to c. Let $R(b)$ be the set of all points accessible from b and let $R^{-1}(b)$ be the set of all points from which b is accessible. If $\overrightarrow{ab}$ is a line of D, then the digraph $D-\overrightarrow{ab}$ has the same set of points as D and has all the lines of D except $\overrightarrow{ab}$. If Z is a directed cycle of D, then by $D-Z$ we mean the digraph obtained from D on removing all the lines of Z.

We assume familiarity with the definition of a *graph* (see [5] or [2]) and a *cycle* of a graph. A *tree* is a connected graph with no cycles. In a *rooted tree*, one of the points, called the *root*, is distinguished. A *directed rooted tree* is obtained from a rooted tree with root b on orienting each of its lines so that it is directed toward the point b. We are now ready to characterize finite functional digraphs.

Theorem 1. A finite digraph D is functional if and only if each of its weak components consists of exactly one directed cycle Z and for each point in Z, the subgraph $R^{-1}(b)$ of the digraph $D-Z$ is a directed rooted tree with root b.

Proof. Let C be a weak component of D, with p points, q lines, and m independent cycles on ignoring the directions of its lines. Then we have $q-p+1=m$. Since C itself is a functional digraph, $p=q$; hence $m=1$. It is then immediately seen that the single independent cycle Z of C must be a directed cycle since each point of Z has out-degree 1. Let b be any point of Z. If b has in-degree 1 then b is an isolated point of $D-Z$. Otherwise, for every point c of $D-Z$ which is joined to b, the line $\overrightarrow{cb}$ is in D. Continuing this argument, we see that the subgraph $R^{-1}(b)$ of the digraph $D-Z$ is a directed rooted tree with root b.

Corollary. Let D be a finite functional digraph. Then D is strong if and only if it consists of exactly one cycle; D is unilateral if and only if it is either strong or consists of exactly one cycle Z together with a path from a point not in Z to a point in Z; finally D is weak if and only if it has exactly one cycle.

We omit the obvious proofs, but depict a strong, unilateral and weak functional digraph in Figures 1(a), 1(b), and 1(c).

We now turn to infinite functional digraphs. Here every directed cycle is of finite length. It is also clear that if c is accessible from b then there is a unique path from b to c and the distance $d(b, c)$ is finite. The concept of a directed rooted tree carries over immediately to infinite digraphs.

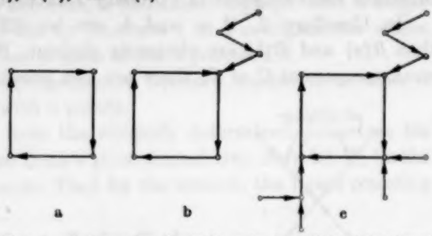


Fig. 1. a) strong functional digraph; b) unilateral functional digraph; c) a weak functional digraph

In proving Theorem 2, it is convenient to have the concept of a semipath in a digraph. A *semipath* of D joining points a_0 and a_n consists of distinct points $a_0, a_1, a_2, \ldots, a_n$ together with n directed lines: exactly one of the lines $\overrightarrow{a_0 a_1}$ or $\overrightarrow{a_1 a_0}$; either $\overrightarrow{a_1 a_2}$ or $\overrightarrow{a_2 a_1}$; ...; and either $\overrightarrow{a_{n-1} a_n}$ or $\overrightarrow{a_n a_{n-1}}$. The concepts of strong, unilateral, weak, and disconnected digraphs extends at once to infinite digraphs. Clearly a digraph (finite or infinite) is weak if and only if there is a semipath between every pair of points.

Theorem 2. An infinite digraph D is functional if and only if the following three conditions hold for each infinite weak component C:

1. C contains at most one cycle.

2. If C has exactly one cycle Z, then for each point b in Z, the subgraph $\langle R^{-1}(b)\rangle$ of $C - Z$ is a directed rooted tree and at least one of these trees is infinite.

3. If C has no cycles, then for each point in C, $\langle R(b)\rangle$ is a ray and $\langle R^{-1}(b)\rangle$ is a directed rooted tree (finite or infinite).

Proof. 1. Assume C contains two cycles Z_1 and Z_2. Since C is functional, there is no semipath joining a point of Z_1 with a point of Z_2, contradicting the hypothesis that C is weak.

2. If c is any point of $C - Z$ from which b is accessible, then $\langle R(c)\rangle$ in $C - Z$ is a unique directed path from c to b. This shows that $\langle R^{-1}(b)\rangle$ in $C - Z$ is a directed rooted tree with root b. At least one of these directed rooted trees must be infinite since C is infinite and Z is finite.

3. That $\langle R(b)\rangle$ is a ray follows at once from the hypotheses that C is infinite, has no cycles, and is functional. We see that $\langle R^{-1}(b)\rangle$ is a directed rooted tree exactly as in the preceding step.

Corollary 1. Let D be a weak infinite functional digraph.

If D has a cycle, then for all points b in D, $R(b)$ is finite.

If D has no cycles, then for all points a and b, there exists a unique point c such that $R(c) = R(a) \cap R(b)$.

Corollary 2. For all points a and b of an arbitrary infinite functional digraph D, $R(a)$ and $R(b)$ are disjoint or differ by only finitely many elements.

We omit the proof of Corollary 1 since it offers no difficulty. However, we illustrate the two parts of Corollary 1 in Figures 2(a) and 2(b) respectively.

In Corollary 2, if a and b are in different weak components of D then $R(a)$ and $R(b)$ are obviously disjoint. But if a and b are in the same weak component C of D, there are two possibilities. If C has a cycle Z, then by statement 1 of Corollary 1, both $R(a)$ and $R(b)$ are finite, so their symmetric difference must be finite. If C has no cycle, then statement 2 of Corollary 1 applies and here the symmetric difference of $R(a)$ and $R(b)$ is the difference between the unique $a \to c$ path (i.e., the path from a to c) and the unique $b \to c$ path in C. Both of these paths are finite and obviously their difference is equal to their union less the single point c, which is necessarily finite.

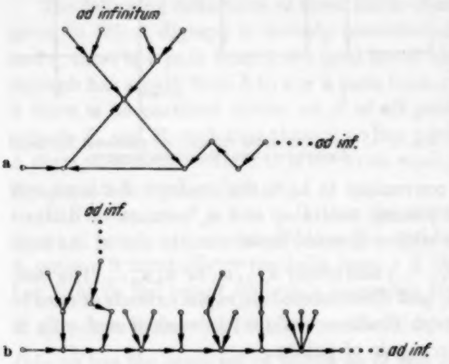


Fig. 2 a) An infinite weak functional digraph containing a cycle; b) an infinite weak functional digraph with no cycles

3. The Number of Functional Digraphs

Two digraphs are *isomorphic* if there is a one-to-one correspondence between their sets of points which preserves their directed lines. To enumerate all finite functional digraphs (up to isomorphism) with a given number of points, we develop the counting series or generating function for these configurations in three steps by enumerating:

1. functional digraphs with a single cycle of specified length k;
2. functional digraphs with any positive number n_k of cycles of length k — these have n_k weak components;
3. arbitrary finite functional digraphs with n_2 cycles of length 2, n_3 cycles of length 3, etc.

These results are obtained by successive applications of Pólya's Enumeration Theorem, which is clearly explained as the Hauptsatz of [5] and also is included in § 2 of [2]; for completeness we include its statement here without the definitions leading up to it.

Pólya's Theorem. The configuration counting series $F(x)$ is obtained by substituting the figure counting series $f(x)$ into the cycle index $Z(\Gamma)$ of the configuration group Γ. Symbolically,

$$F(x) = Z(\Gamma, f(x)) . \tag{1}$$

This theorem reduces the problem of finding the configuration counting series to the determination of the figure counting series and the cycle index of the configuration group.

Step 1. Functional digraphs with a single cycle of length k

In the framework of Pólya's theorem, a functional digraph with a single cycle Z of length k may be regarded by Theorem 1 as a configuration whose figures are the directed rooted trees at each of the k points of Z.

Remark. There is a $1-1$ correspondence between rooted trees with n points and directed rooted trees with n points.

This remark follows at once from the uniquely determined procedure for constructing a directed rooted tree from a given rooted tree. Now let T_n be the number of rooted trees with n points. Then by the remark, the figure counting series for Step 1 is:

$$T(x) = T_1 x + T_2 x^2 + T_3 x^3 + \cdots. \tag{2}$$

Cayley has shown that

$$T(x) = x(1-x)^{-T_1}(1-x^2)^{-T_2}(1-x^3)^{-T_3}\ldots, \tag{3}$$

and Pólya [5], using his theorem, obtained the equivalent functional equation

$$T(x) = x \exp \sum_{r=1}^{\infty} T(x^r)/r. \tag{3'}$$

Using either (3) or (3′), one may compute the values of T_n, and Riordan [8], p. 138, has calculated these values to $n = 26$. The first few terms in this series are:

$$\begin{aligned} T(x) = {} & x + x^2 + 2x^3 + 4x^4 + 9x^5 + 20x^6 + 48x^7 + 115x^8 + \\ & + 286x^9 + 719x^{10} + 1842x^{11} + 4766x^{12} + \cdots. \end{aligned} \tag{2'}$$

In addition, Prins has drawn all the trees with < 12 points; the coefficients in (2′) check with the number of dissimilar points in these trees. Diagrams of all trees with < 10 points appear in [4].

Since these k rooted trees are located at the points of the directed cycle Z, the configuration group for Step 1 is C_k, the cyclic group of degree k and order k. The cycle index of C_n is given by:

$$Z(C_n) = \frac{1}{n} \sum_{d|n} \varphi(d) f_d^{n/d}, \tag{4}$$

where the letters $f_1, f_2, \ldots, f_n$ are indeterminates, the sum is taken over all divisors d of n, and $\varphi(d)$ is the Euler function.

Applying Pólya's theorem, we find that the configuration counting series for Step 1 is

$$Z(C_k, T(x)) = \frac{1}{k} \sum_{d|k} \varphi(d)\,(T(x^d))^{k/d}. \tag{5}$$

The coefficient of x^p in (5) is the number of weak functional digraphs with p points whose cycle has length k.

Step 2. Functional digraphs with n_k cycles of length k

It is convenient in this step to regard a figure as a functional digraph with a single cycle of length k. Then the figure counting series for Step 2 is the configuration counting series for Step 1, which is given by (5). Since there are now n_k figures in each configuration and it makes no difference whatsoever

where each figure is located, the configuration group for Step 2 is the symmetric group S_{n_k} of degree n_k. The cycle index of S_n is given by

$$(6) \qquad Z(S_n) = \frac{1}{n!} \sum_{(j)} \frac{n!}{1^{j_1}\, j_1! \cdots n^{j_n}\, j_n!} f_1^{j_1} \cdots f_n^{j_n}$$

where the sum is taken over all partitions $(j) = (j_1, j_2, \ldots, j_n)$ of n such that

$$(7) \qquad 1j_1 + 2j_2 + \cdots + nj_n = n\,.$$

For completeness and later use, we define $Z(S_0) = 1$.

Combining the observations and applying Pólya's Theorem, we see that the configuration series for Step 2 is

$$(8) \qquad Z(S_{n_k}, Z(C_k, T(x)))\,.$$

Pólya [5] has developed the „Gruppenkranz" $A[B]$ of two permutation groups A and B, called their *composition* in [3], and by one of his results, we may rewrite (8) in the form

$$(8') \qquad Z(S_{n_k}[C_k], T(x))\,.$$

The coefficient of x^p in (8) or (8′) is the number of functional digraphs with n_k cycles of length k having p points.

Step 3. Arbitrary finite functional digraphs

Let D be a finite functional digraph with n_k cycles of length k for $k = 2, 3, \ldots$. Let v_p be the number of functional digraphs with p points and let the desired configuration series be

$$(9) \qquad v(x) = v_2 x^2 + v_3 x^3 + \cdots\,.$$

Then it is clear from Step 2 that $v(x)$ may be written

$$(10) \qquad v(x) = \sum_{\text{all } n_k = 0}^{\infty} \prod_{k=2}^{\infty} Z\bigl(S_{n_k}, Z(C_k, T(x))\bigr)\,.$$

On interchanging the sum and product symbols we obtain

$$(11) \qquad v(x) = \prod_{k=2}^{\infty} \sum_{r=0}^{\infty} Z\bigl(S_r, Z(C_k, T(x))\bigr)\,.$$

But there is a well known identity for any function $a(x)$:

$$(12) \qquad \sum_{r=0}^{\infty} Z(S_r, a(x)) = \exp \sum_{m=1}^{\infty} \frac{1}{m} a(x^m)\,.$$

Using (12), equation (11) may be rewritten

$$\begin{aligned} v(x) &= \prod_{k=2}^{\infty} \exp \sum_{m=1}^{\infty} \frac{1}{m} Z(C_k, T(x^m)) \\ &= \exp \sum_{k=2}^{\infty} \sum_{m=1}^{\infty} \frac{1}{m} Z(C_k, T(x^m)) \\ &= \exp \sum_{m=1}^{\infty} \frac{1}{m} \sum_{k=2}^{\infty} Z(C_k, T(x^m))\,. \end{aligned}$$

If one could find an explicit formulation for the sum

$$\sum_{k=2}^{\infty} Z(C_k, a(x))$$

for any function $a(x)$, corresponding to (12), then the formula for $v(x)$ could be reduced further.

4. Functional digraphs with various kinds of connectedness

Obviously, every strong digraph is unilateral and every unilateral digraph is weak. In order to divide all digraphs into disjoint connectedness categories, let U_3 be the set of all strong digraphs, U_2 the set of all unilateral digraphs which are not strong, and U_1 contain all other weak digraphs; and let U_0 be the set of all disconnected digraphs.

We wish to enumerate the functional digraphs which are in each of these classes U_k. Let $v_k(x)$ be the counting series for the functional digraphs which are in U_k. Since these connectedness categories are disjoint,

$$v(x) = v_3(x) + v_2(x) + v_1(x) + v_0(x)\,. \tag{13}$$

The counting series for strong functional digraphs is immediately obtained from the fact that such a digraph consists of exactly one directed cycle. Hence $v_3(x) = x^2 + x^3 + x^4 + \cdots$, or

$$v_3(x) = \frac{x^2}{1-x}\,. \tag{14}$$

A functional digraph is *strictly unilateral*, i.e., in U_2, if it contains exactly one weak component C of the following form. There is at least one point of C which doesn't lie in its directed cycle, Z, and the set of all such points generates a path. Thus only one of the points of Z has a rooted tree containing more than one point and this tree consists of a single path. From these considerations, we see that the number of strictly unilateral functional digraphs is given by

$$v_2(x) = x^3 + 2x^4 + 3x^5 + \cdots = \frac{x^3}{(1-x)^2}\,. \tag{15}$$

The weak functional digraphs are enumerated from $v(x)$ using the formula from [2] which expresses the number of connected graphs of any given kind in terms of the total number of graphs of that kind. This leads to the equation

$$\sum_{n=1}^{\infty} \frac{1}{n}\,[v_1(x^n) + v_2(x^n) + v_3(x^n)] = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n}\,(v(x))^n. \tag{16}$$

Using equations (11), (14), and (15) for $v(x)$, $v_3(x)$ and $v_2(x)$, equation (16) becomes

$$\sum_{n=1}^{\infty} \frac{1}{x}\,v_1(x^n) = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n}\,(v(x))^n - \sum_{n=1}^{\infty} \frac{x^{2n}}{n(1-x^n)^2}\,, \tag{16'}$$

which serves to determine $v_1(x)$. Then an application of (13) gives $v_0(x)$.

5. The Number of Types of Functions

The principal conceptual difference between a function and a functional digraph is that a functional digraph has no lines of the form $\overrightarrow{bb}$. Thus, the counting series of the preceding section serves to enumerate the isomorphism types of the class of all functions which are *fixed-point-free*. In order to count all functions, we need to modify the above formulas by allowing for slings, or cycles of length 1. The structural formulas of section 2 will still hold, with the stipulation that the word cycle is extended to admit cycles of length 1,

We conclude by indicating without further proof the formula corresponding to (11) for the total number of types of functions from a finite set into itself, i.e., functions on a finite set:

$$\prod_{k=1}^{\infty} \sum_{r=0}^{\infty} Z\left(S_r, Z\left(C_k, T(x)\right)\right).$$

In this formula the coefficient of x^n is the formula "$fcn(n)$" of Davis [1].

References

[1] Davis, R. L.: The number of structures of finite relations. Proc. Amer. math. Soc. 4, 486—495 (1953). — [2] Harary, F.: The number of linear, directed, rooted and connected graphs. Trans. Amer. math. Soc. 78, 445—463 (1955). — [3] Harary, F.: On the group of the composition of two graphs. Duke math. J. 26, 29—34 (1959). — [4] Harary, F., and G. Prins: The number of homeomorphically irreducible trees, and other species, Acta math. 101, 141—162 (1959). — [5] König, D.: Theorie der endlichen und unendlichen Graphen. Leipzig 1936. — [6] Myhill, J.: Recursive digraphs, splinters and cylinders. Math. Annalen 138, 211—218 (1959). — [7] Pólya, G.: Kombinatorische Anzahlbestimmungen für Gruppen, Graphen und chemische Verbindungen. Acta math. 68, 145—254 (1937). — [8] Riordan, J.: An introduction to combinatorial analysis. New York 1958.

(*Received February 20, 1959*)

MYHILL, J.
Math. Annalen 138, 211—218 (1959)

Recursive Digraphs, Splinters and Cylinders*

By

JOHN MYHILL in Princeton, N. J.

Connections between recursion theory and graph theory appear in DEKKER and MYHILL [5] and are implicit in KLEENE [7]. A third overlap was suggested to the writer by a talk of ULLIAN [10]; § 1 of the present paper presents some of Ullian's results in a graph-theoretic context. § 2 develops some further properties of the recursively enumerable (r. e.) sets discussed by ULLIAN, and § 3 contains a theorem on recursive digraphs of another kind. We presuppose a moderate acquaintance with recursion theory (as contained e.g. in [1]); for graph theory we use the terminology and one or two of the results of HARARY [6].

§ 1. Splinters and Components

With every recursive function f we associate the digraph $G(f)$ having as its points the non-negative integers and as its lines all pairs $(n, f(n))$. (Of course $G(f)$ may contain slings.) To avoid needless repetition we let f be fixed throughout the present section; all graphs considered are subgraphs of $G(f)$. We do not usually distinguish between a graph and its point-set. When we say 'connected' or 'component', 'weakly connected' and 'weak component' will always be understood. A *splinter* (ULLIAN) is a set of the form $\{f^n(a_0)\}$. Graph-theoretically, a splinter is connected, and as such, is included in exactly one component; we call this *the component of the splinter*. In the notation of [6] the splinter $\{f^n(a_0)\}$ is $R(a_0)$.

Theorem 1.1. *Every splinter S is recursive in its component C.*

Proof. If C contains a cycle, S is finite ([6] Theorem 2, Corollary 1), so we can assume C contains no cycle. Given a decision-procedure for C we can decide of an arbitrary element x whether it belongs to S in the following way. First ask if $x \in C$. If not we are through. If x is in C, generate the rays $R(x)$ and $S = R(a_0)$ (notation as in [6]). In a finite number of steps we obtain a unique number c such that $R(c) = [R(x) \cap R(a_0)]$ ([6] Theorem 2, Corollary 1); further when we have obtained such a c we shall know it. Now $x \in S$ or $x \in C - S$ according as $x = c$ or $x \neq c$, q. e. d.

We note that *$C - S$ is r. e.* For it consists of precisely those elements x of C for which the computation of c terminates by yielding a number $x \neq c$.

Define a component to be *cyclic* if it contains a cycle. We prove

*) This work was supported by grants from the National Science Foundation (G 3466) and the Institute for Advanced Study. I wish to thank F. HARARY and D. LACOMBE for ideas which are incorporated into this paper.

Theorem 1.2. *Every cyclic component C_0 is recursive in the union ΣC of all cyclic components.*

Proof. C_0 contains only one cycle ([6], Theorem 2). Let this cycle be $(x_1, \ldots, x_n, x_1)$. For all $y \in \Sigma C$, $R(y)$ is finite ([6], Theorem 2, Corollary 1). If we know that $y \in \Sigma C$, we can find whether or not $y \in C_0$ by generating $R(y)$ and observing whether or not any of the elements $x_1, \ldots, x_n$ appear in it, q. e. d.

Again we have the stronger result that *$\Sigma C - C_0$ is r. e.*

A set A is called a *choice-set* of a class $\mathfrak{A}$ of sets, if A contains exactly one element of each set belonging to $\mathfrak{A}$, and no other elements. We prove

Theorem 1.3. *The class of all cyclic components possesses a r. e. choice-set.*

Proof. The set of all minimal elements of cycles of $G(f)$ is such a set.

Theorem 1.4. *If the class of all components possesses a r. e. choice-set, every component is recursive.*

Proof. Let H be the choice set, C an arbitrary component. Then $C \cap H$ is a singleton, say $\{x_0\}$. Given any element x we can determine whether or not $x \in C$ by generating the component of x till an element of H is obtained; x is or is not in C according as this element is or is not equal to x_0.

Corollary 1. *If all but finitely many components are cyclic, every component is recursive.* (By Theorem 1.3.)

Corollary 2. *No component is simple.*

Proof. If possible, let C be a simple component. Then every other component is r. e. and disjoint from C; hence finite; hence cyclic ([6], Theorem 1). The result now follows from Corollary 1.

The following theorem generalizes Corollary 2.

Theorem 1.5[1]**.** *If C is a non-recursive but recursively enumerable union of components, at most finitely many of which are cyclic, and if D is a r. e. set disjoint from C, then $C \cup D$ is not simple.*

Proof. We assume that $C \cup D$ is simple and show how to decide of an arbitrary number x whether or not it belongs to C; this will yield a contradiction.

For every element x, $R(x)$ either (i) meets C or (ii) meets D or (iii) contains a cycle. For if $R(x)$ satisfied neither (i), (ii) nor (iii), it would be an infinite r. e. subset of $N - (C \cup D)$ (where N is the set of all natural numbers). For an arbitrary element x, generate $R(x)$ till either (i) we obtain an element of C or (ii) we obtain an element of D or (iii) we obtain a cycle. In case (i) x is in C; in case (ii) it is not; in case (iii) it is in C or not according as the cycle occurring in $R(x)$ is or is not one of the finite number of cycles occurring in C.

Corollary. *If C is a non-recursive component and D is a r. e. set disjoint from C, then $C \cup D$ is not simple.*

For splinters we have

Theorem 1.6. (Ullian [10]). *If S is a non-recursive splinter and D is a r. e. set disjoint from S, then $S \cup D$ is not simple.* A fortiori, *S itself is not simple.*

[1]) This theorem is due to D. Lacombe.

Proof. Again we assume that $S \cup D$ is simple and obtain on this hypothesis a decision-procedure for S, contradicting the assumption that S is not recursive. Let C be the component of S; then for every element x either (i) $x \in S$ or (ii) $x \in C - S$ or (iii) $R(x)$ meets D or (iv) $R(x)$ contains a cycle. For if (i) and (ii) were false x and therefore $R(x)$ would lie outside C; if in addition (iii) were false $R(x)$ would be a r. e. subset of $N - (S \cup D)$; and if finally (iv) were false $S \cup D$ would not be simple. For an arbitrary element x, generate S, $C - S$, $R(x)$ and D till one of the alternatives (i)—(iv) is realized. ($C - S$ can be effectively generated by the remark following Theorem 1.1.) If (i) or (ii) is realized, we know whether $x \in S$. If (iii) is realized x cannot be in S, because $x \in S$ would imply $R(x) \subset S$ and so $R(x) \cap D = \theta$. If finally (iv) is realized x cannot be in S either, because $R(x)$ is infinite for each $x \in S$ (otherwise S would be finite). Thus the decision-problem for S is solvable: contradiction.

§ 2. Cylinders and *m*-*n*-Splinters

Let N be the set of natural numbers, and let j be a recursive function (assumed fixed in what follows) mapping N^2 one-one on N. By $Cy(A)$ for any $A \subset N$ we denote the set $j(A \times N)$. A set M is called a *cylinder* if $M \cong Cy(A)$ for some non-empty r. e. set A. Hence $\cong$ means isomorphism in the sense of [8], p. 106.

Theorem 2.1. *Every cylinder is a splinter*[2]).

Proof. Since a set isomorphic to a splinter is itself a splinter, it suffices to prove the theorem for sets $Cy(A) = j(A \times N)$ for A non-empty and r. e. Also we can assume that A is infinite, for if A were finite $Cy(A)$ would be recursive and trivially a splinter. Let then $A = \{d_i\}$ where d_i is a one-one recursive function of i. We have $Cy(A) = M_0 + M_1$, where

$$M_0 = \{j(d_x, y) | x \leqq y\}$$
$$M_1 = \{j(d_x, y) | x > y\}\,.$$

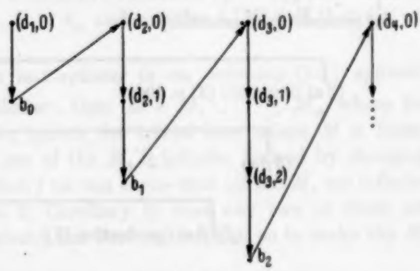


Fig. 1

M_0 is an infinite recursive set and can therefore be enumerated by a strictly monotone recursive function b_i. Every element of $Cy(A)$ appears exactly once in Fig. 1, where for legibility the j's have been omitted.

It now suffices to construct a recursive function f such that the arrow from any number x in Fig. 1 goes to the number $f(x)$. For then $Cy(A) = \{f^n j(d_1, 0)\}$. But such a function f is evidently obtained if we let $k(x)$, $l(x)$ be recursive functions satisfying

$$j(k(x), l(x)) = x$$

[2]) This is a simplification of a proof due to D. Lacombe.

and define

$$f(x) = \begin{cases} b_{l(x)}, \text{ if } k(x) = d_{l(x)+1} \\ j(d_{n+2}, 0), \text{ if } x = b_n \\ j(k(x), l(x)+1), \text{ otherwise.} \end{cases}$$

Corollary 1. ***For every non-empty r. e. set A, there is a splinter S such that A is one-one reducible to S ([9], p. 297) and S is strongly (many-one) reducible to A. (Hence every degree of many-one reducibility of r. e. sets is represented by a splinter.)***

Proof.

$$X \in A \leftrightarrow j(x, 0) \in Cy(A)$$
$$X \in Cy(A) \leftrightarrow k(x) \in A.$$

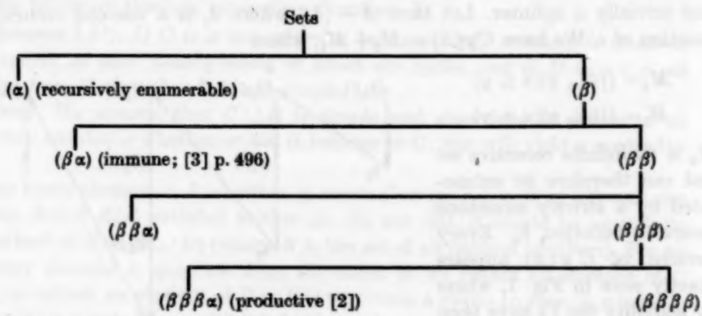
Corollary 2. ***Every creative set is a splinter.***

Proof. Let A be creative. By Corollary 1 A is one-one reducible to some splinter S. By [8] Theorem 16(a), S is one-one reducible to A. By [8] Theorem 18 $A \cong S$ and so is a splinter.

Corollary 3. ***There exist mesoic splinters ([3] p. 499); in fact for every simple set A, $Cy(A)$ is a mesoic splinter.***

Proof. $Cy(A)$ is a splinter by Theorem 2.1, and it is mesoic by the argument of [11] Theorem 3.

Remark[3]). The traditional classification of r. e. sets ([9], [3], [11]) is by the nature of their complements. In the most recent form of this classification (USPENSKI) arbitrary sets are classified according to the following scheme



Here $(\beta\beta\alpha)$ is the class of all sets $M_1 \cup M_2$ where $M_1 \cap M_2 = \theta$, M_1 is infinite and r. e. and M_2 is immune; and $(\beta\beta\beta)$ is the class of all sets M which are not r. e., which contain an infinite r. e. subset M_1 and which are such that for every infinite r. e. set $M_2 \subset M$ there is an infinite r. e. set $M_3 \subset M - M_2$.

[3]) To avoid repetitiousness we make the conventions that (*a*) 'set' will mean 'set of non-negative integers' for the remainder of the paper and (*b*) *for the duration of the present remark* we shall ignore finite sets and sets with finite complements.

$(\beta\beta\beta)$ is then subdivided into $(\beta\beta\beta\alpha)$ and $(\beta\beta\beta\beta)$ according as (an index of) such an M_3 can or cannot be effectively found from (an index of) M_2. (This definition of 'productive' is equivalent to the usual one by [11], pp. 168—169). Recursively enumerable sets M are now further classified into five types according as $N-M$ belongs to (α), $(\beta\alpha)$, $(\beta\beta\alpha)$, $(\beta\beta\beta\alpha)$ or $(\beta\beta\beta\beta)$; there exist r. e. sets of all five types by [9] pp. 296—298; [3] p. 500; [11] p. 164. The class of all splinters is unusual in that it is the only (recursively invariant) class of r. e. sets appearing in the literature which is defined without reference to complements. It is natural to ask to which of the five types (α), $(\beta\alpha)$, $(\beta\beta\alpha)$, $(\beta\beta\beta\alpha)$, $(\beta\beta\beta\beta)$ the set $N-M$ may belong if M is a splinter. But this is already answered by the preceding results of this paper; (α), $(\beta\beta\beta\alpha)$ and $(\beta\beta\beta\beta)$ are possible by Corollaries 2—3 of Theorem 2.6, while $(\beta\alpha)$ and $(\beta\beta\alpha)$ are impossible by Theorem 1.6. Indeed this observation is already contained in [10]. Recently a finer classification of r. e. sets has been made possible by the theory of *recursive equivalence types* ([4]). Indeed a r. e. set is characterized to within an isomorphism by the recursive equivalence type of its complement ([4], Theorem 6), and it is easily proved using Theorem 24 of [4] that a r. e. set M is a cylinder if and only if the recursive equivalence type A of $N-M$ satisfies the equation $A = 2A$. If then we could prove the converse of Theorem 2.1 we would have a complete characterization of splinters, of a very simple kind, in terms of the recursive equivalence types of their complements. There is therefore a strong motivation for trying to prove (or disprove) that every (infinite) splinter is a cylinder; however we have not been able to do this[4]), and we do not consider the problem a trivial one.

A natural generalization of the idea of a splinter is that of an *m-n-splinter* $(m, n \geq 1)$; a set M is called an m-n-splinter if there exist m non-negative integers $a_1, \ldots, a_m$ and n recursive functions $f_1, \ldots, f_n$ such that M is the least set which contains $a_1, \ldots, a_m$ and satisfies $f_i(M) \subset M$ $(1 \leq i \leq n)$. We prove the

Theorem 2.2. *For* $m \geq 1$, *an* m-1-*splinter is an ordinary* (1-1) *splinter.*

Proof. Let M be an m-1-splinter; then $M = M_1 \cup \cdots \cup M_m$ where for $1 \leq s \leq m$ M_s is $\{f^i(a_s)\}$. We can ignore the trivial case where M is finite, i.e. we can assume that at least one of the M_s is infinite. Indeed by changing finitely many values of the function f we can ensue that *all* the M_s are infinite, and further (using [6] Theorem 2, Corollary 2) that any two of them are disjoint. Henceforth we assume that f has been so modified so to make the M_s infinite and disjoint.

For $1 \leq s \leq m$ define $m_s(0) = a_s$; $m_s(x+1) =$ the first term of $\{f^i(a_s)\}$ which (follows and) exceeds $m_s(x)$; $M_s^* = \{m_s(i)\}$. Then $M^* = M_1^* \cup \cdots \cup M_m^*$ is a recursive set, and for each $x \in M^*$ we can effectively find numbers s, i for which $x = m_s(i)$. In what follows we shall assume that $m = 4$; the modifications required for other m are obvious.

[4]) D. Lacombe has proved that every (infinite) splinter *of a one-one recursive function* is a cylinder, and also that every (infinite) *component* of $G(f)$ is a cylinder for *any* recursive f.

Let $f^*(x) = f(x)$ for $x \notin M^*$; and for any $x \in M^*$, say $x = m_s(i)$, let $f^*(x)$ have the value given in the following table.

Table

$s \backslash i$	0	1	$2n+2$	$2n+3$
1	$f(x)$	$m_2(0)$	$f(x)$	$fm_2(2n+2)$
2	$f(x)$	$m_3(0)$	$fm_1(2n+1)$	$fm_3(2n+2)$
3	$f(x)$	$m_4(0)$	$fm_2(2n+1)$	$fm_4(2n+2)$
4	$f(x)$	$f(x)$	$fm_3(2n+1)$	$f(x)$

With f^* so defined, the sequence $\{f_i^* m_1(0)\}$ will run through $M = M_1 \cup M_2 \cup M_3 \cup M_4$ in the order shown in Fig. 2.
Thus M is identical with the splinter $\{f_i^* m_1(0)\}$, q. e. d.

Theorem 2.3. *Every* r. e. *set is a* 1-2-*splinter.*

Proof. We need first the following lemma, which despite its elementary character seems to have escaped notice in the literature.

Every infinite r. e. set is the union of a creative set and an infinite recursive set.

Proof of the lemma. Let M be an infinite r. e. set and let M_0 be an infinite recursive subset of M ([9], p. 291). Let $m(x)$ be a one-one recursive function enumerating M_0, and let C be creative. Then $m(C) \subset M_0$ is creative (cf. [8], Theorem 5), and $M - M_0$ is a r. e. set disjoint from $m(C)$. Thus $M_1 = m(C) + (M - M_0)$ is creative ([9], p. 296), M_0 is infinite and recursive, and $M = M_0 + M_1$, q. e. d.

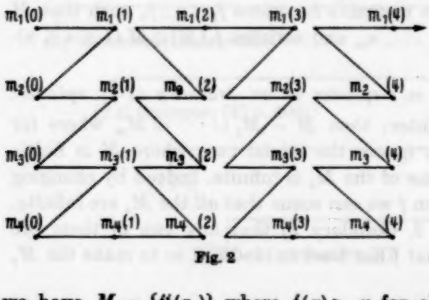


Fig. 2

Now we can complete the proof of Theorem 2.3. Let M be r. e.; without loss of generality we can assume that M is infinite. By the lemma, $M = M_0 + M_1$, where M_0 is infinite and recursive and M_1 is creative. Then if a_0 is the least element of M_0 and if we define

$$f(x) = (\mu y)(y \in M_0 \,\&\, y > x)$$

we have $M_0 = \{f^i(a_0)\}$ where $f(x) > x$ for all x. Further by Corollary 2 of Theorem 2.1 M_1 is a splinter, say $M_1 = \{g^i(b_0)\}$. Without loss of generality we can suppose $b_0 \notin M_0$. Now let h_0, h_1 be two recursive functions defined by

$$h_0(x) = \begin{cases} g^i(b_0), \text{ if } x = f^i(a_0) \\ a_0, \text{ if } x \notin M_0 \end{cases}$$

$$h_1(x) = \begin{cases} f(x), \text{ if } x \in M_0 \\ a_0, \text{ if } x \notin M_0 \,. \end{cases}$$

$h_0(x)$ is always in $M_1 \cup \{a_0\} \subset M$, and $h_1(x)$ is always in $M_0 \subset M$. Thus M is closed under h_0 and h_1. We claim that it is the *least* set containing b_0 which is closed under these two functions. Indeed, if a set U contains b_0 and is closed under h_0 and h_1, it contains a_0 (by the definition of h_0 and the fact that $b_0 \notin M_0$; hence it includes M_0 (by the definition of h_1) and M_1 (by the definition of h_0). Thus M is a 2-1 splinter, q. e. d.

According to Theorem 2.2 the notion of m-n-splinter coincides with the ordinary notion of splinter for $m \geqq 1$, $n = 1$; and according to Theorem 2.3 it coincides with the notion of r. e. set for $m \geqq 1$, $n \geqq 2$.

§ 3. Universal Digraphs

Now we turn to digraphs of a different kind. A collection $\mathfrak{R}$ of digraphs[5]) is said to contain a *universal digraph* U, if $U \in \mathfrak{R}$ and if every $G \in \mathfrak{R}$ is isomorphic to a subgraph of U. We claim that there exists a universal digraph for the collection of all countable digraphs.

We define the digraph U as follows. By a *labelled digraph* with n points we shall mean the empty digraph if $n = 0$; a digraph whose points are the number $(1, \ldots, n)$, if $0 < n < \infty$; or a digraph whose points are all the positive integers, if $n = \infty$. G_1 with n_1 points $(1, \ldots, n_1)$ is called a *labelled subgraph* of G_2 with $n_2 \geqq n_1$ points if every $p, q \leqq n_1$ are joined in G_1 just in case they are joined in G_2.

The points of U are simply *all finite labelled digraphs*. G_1 is joined to G_2 in U if and only if one of them is a subgraph of the other and if, where n_1 is the number of points in G_1 and n_2 the number of points in G_2, n_1 is joined to n_2 in G_1 (if $G_2 \subset G_1$) or in G_2 (if $G_1 \subset G_2$).

Now let G be any labelled digraph with s points. For $n = 1, 2, 3, \ldots, s$ let G_n be the subgraph of G generated by $1, \ldots, n$. We claim that the mapping φ defined by

$$\varphi(n) = G_n$$

maps G isomorphically into U; i.e. k is joined to m in G if and only if G_k is joined to G_m in U. Indeed if $k \leqq m$, $G_k \subset G_m$ and G_k is joined to G_m in U if and only if k is joined to m in G_m, i.e. if and only if k is joined to m in G. On the other hand, if $k > m$, $G_m \subset G_k$ and G_k is joined to G_m in U if and only if k is joined to m in G_k, i.e. once more if and only if k is joined to m in G. Thus every labelled digraph G (hence every countable digraph) is isomorphic to a subgraph of U. Finally, U itself is clearly countable, since there are only countably many finite labelled digraphs. U is also recursive in the sense that the relation 'm is joined to n in U' is recursive, on any natural Gödel-numbering of finite labelled digraphs.

A less trivial universal digraph for the collection of all countable digraphs is provided by the following example. Let $\{W_i\}$ be the canonical enumeration of r. e. sets ([3] p. 498, or [8] p. 99; the notation $\{\omega_i\}$ is used there), and let U_1 be the labelled digraph of the relation '$x \in W_i$'. We prove

---

[5]) For the remainder of the paper, digraphs are allowed to contain slings and isolated points, but no multiple lines.

Theorem 3.1. *(a) U_1 is a universal countable digraph. Moreover (b) if we define $R_n(x, y)$ to mean $j(x, y) \in W_n$, there exists a recursive function h such that (the digraph of) R_n is isomorphic to the subgraph $\langle W_{h(n)} \rangle$ of U_1.*

Proof. (b) implies (a). For the relation 'x is joined to y in U' is a recursive and a fortiori a r. e. relation, say $R_n(x, y)$. Moreover every countable digraph is isomorphic to a subgraph of U, which is in turn by (b) isomorphic to the subgraph $\langle W_{h(n)} \rangle$ of U_1.

Let $\{q_n(x)\} = \{q(n, x)\}$ be the canonical enumeration of partial recursive functions ([8], p. 99). Let $a(u, v, z)$ be a recursive function such that always

$$W_{a(u,v,z)} = \{q_u(x) | R_v(x, z)\} .$$

We can assume that a maps N^3 one-one into N. For fixed n_0 there exists by the recursion theorem a number e such that always

$$q_e(z) = a(e, n_0, z) .$$

But can be found effectively from n_0 (cf. the proofs of Theorems 7.4—5 of [1], Chapter 10); there exists therefore a recursive function $e(n)$ such that always

$$q(e(n), z) = a(e(n), n, z) .$$

We have

$$\begin{aligned} q(e(n), x) \in W_{q(e(n), y)} &\leftrightarrow q(e(n, x) \in W_{a(e(n), n, y)} \\ &\leftrightarrow q(e(n), x) \in \{q_{e(n)}(m) | R_n(m, y)\} \\ &\leftrightarrow R_n(x, y) . \end{aligned}$$

(The last step is correct because $q_{e(n)}$ is one-one for fixed n.) Thus R_n is isomorphic, under the mapping

$$x \to q(e(n), x)$$

to the subgraph of U generated by

$$H_n = \{q(e(n), x) | (\exists y) (R_n(x, y) \quad \text{or} \quad R_n(y, x))\}$$

and (b) is proved if we take $h(n)$ as a Gödel-number of H_n.

References

[1] Davis, M.: Computability and Unsolvability. New York 1958. — [2] Dekker, J. C. E.: Productive sets. Trans. Amer. math. Soc. 78, 129—149 (1955). — [3] Dekker, J. C. E.: Two notes on recursively enumerable sets. Proc. Amer. math. Soc. 4, 495—501 (1953). — [4] Dekker, J. C. E., and J. Myhill: Recursive equivalence types, to be published by the University of California Publications in Mathematics. — [5] Dekker, J. C. E., and J. Myhill: Retraceable sets. Canad. J. Math. 10, 357—373 (1958). — [6] Harary, F.: The number of functional digraphs. Math. Ann. 138, 203—210 (1959).— [7] Kleene, S. C.: Recursive functions and intuitionistic mathematics. Proc. Int. Congress of Mathematicians. vol. 1, pp. 679—685. Cambridge, Mass., U.S.A. (1950; published 1952). — [8] Myhill, J.: Creative sets. Z. math. Log. u. Grundlagen d. Math. 1, 97—108 (1955). — [9] Post, E. L.: Recursively enumerable sets and their decision-problems. Bull. Amer. math. Soc. 50, 284—316 (1944). — [10] Ullian, J.: Splinters of recursive functions (abstract of a paper read at a meeting of the Association for Symbolic Logic, December 1957). J. Symb. Log. 23, 107—108 (1958). — [11] Uspenski, V. A.: Neskol'ko zamečanij o perečislimykh množestvakh. Z. math. Log. u. Grundlagen d. Math. 3, 157—170 (1957).

(Received February 20, 1959)

Foster, A. L.
Math. Annalen 138, 219—238 (1959)

On the Imbeddability of Universal Algebras in Relation to their Identities. I

By

Alfred L. Foster in Berkeley (Cal.)

1. Introduction

We seek to establish a foundational framework within which a class of questions related to algebraic imbeddability may be formulated and explored on a universal algebra level. Decisively bound to the anatomy of imbeddability of algebras are certain elementary restrictive relationships between the sets of identities (and/or equations) of the respective algebras involved. We shall be concerned with a definition and analysis of these relationships. Within this general framework, and with strong emphasis on the identities of algebras, several characterizations of two basic imbeddability questions are obtained, and related questions on "universal" imbeddability are explored. As significant instances of the latter notion the singular role played by primal algebras is exhibited and further, a more general class of "categorical" algebras is characterized.

2. Some preliminaries

Let S and S' be arbitrary but fixed species, with S a subspecies of S' (S' an overspecies of S):

(2.1) $$S \subseteq S'\,; \quad S = (n_1, n_2, \ldots)\,, \quad S' = (n_1, n_2, \ldots, n'_1, \ldots)\,.$$

Here the n_i and n'_j are positive rational integers.

We shall reserve the letters $\xi, \eta, \zeta, \xi_1, \xi_2, \ldots, \eta_1, \eta_2, \ldots$ to denote indeterminante-symbols, and $\sigma_1, \sigma_2, \ldots, \sigma'_1, \ldots$ to denote primitive operation-symbols, where σ_i is n_i-ary, $\sigma_i = \sigma_i(\xi_1, \xi_2, \ldots, \xi_{n_i})$ and σ'_j is n'_j-ary.

An "S-algebra" $\mathfrak{A} = (A, \sigma_1, \sigma_2, \ldots)$ is simply an algebra of species S. If $\mathfrak{B} = (B, \sigma_1, \sigma_2, \ldots)$ is also an S-algebra, the primitive operations σ_i of the algebra $\mathfrak{A}$ — read $\sigma_i\,(\mathrm{mod}\,\mathfrak{A})$ — will of course in general be quite unrelated to the corresponding operation $\sigma_i\,(\mathrm{mod}\,\mathfrak{B})$.

We shall write

(2.2) $$\mathfrak{A} < \mathfrak{A}'$$

to signify that the S-algebra $\mathfrak{A}$ is imbedded in the S'-algebra

$$\mathfrak{A}' = (A', \sigma_1, \sigma_2, \ldots, \sigma'_1, \ldots)\,,$$

(2.2.1) $A \subseteq A'$, $\sigma_i(\xi, \ldots) \in A$ for $\xi, \ldots \in A$ and σ_i any S-operation.

The relation (2.2) is also read: $\mathfrak{A}$ is an "S-modul" of $\mathfrak{A}'$.

If $\mathfrak{A} < \mathfrak{A}'$ and $A = A'$ we also write

(2.3) $$\mathfrak{A} \subset< \mathfrak{A}'$$

read: $\mathfrak{A}'$ is an "S'-conversion" of $\mathfrak{A}$, also: $\mathfrak{A}$ is $\mathfrak{A}'$ *qua* S-algebra.

If $\mathfrak{A} < \mathfrak{A}'$ and $S = S'$ we have the case: $\mathfrak{A}$ is a subalgebra of $\mathfrak{A}'$. Unless otherwise specified we hereafter assume proper inclusion, $S \subset S'$.

We write $|\mathfrak{A}'|$ for the class of all identities of $\mathfrak{A}'$, and $|\mathfrak{A}'|_S$ for the subclass of all S-identities (i.e., in which no σ'_j occur),

(2.4) $$|\mathfrak{A}'|_S \subseteq |\mathfrak{A}'| .$$

An "S'-expression" $\Phi'(\xi, \ldots)$ is an indeterminate-symbol or any composition of indeterminate-symbols via the primitive S'-operations $\sigma_1, \sigma_2, \ldots, \sigma'_1, \ldots$. Let

(2.5) $$E(S') = E' = \{\ldots, \Phi', \ldots\}$$

denote the totality of S'-expressions, and write

(2.6) $$\mathcal{E}(S') = \mathcal{E}' = (E', \sigma_1, \sigma_2, \ldots, \sigma'_1, \ldots)$$

for the "basic" S'-algebra, in which $\sigma_1, \sigma_2, \ldots, \sigma'_1, \ldots$ are applied formally in E'. With $E(S) = E$ and $\mathcal{E}(S) = \mathcal{E} = (E, \sigma_1, \sigma_2, \ldots)$ similarly defined, clearly

(2.7) $$\mathcal{E} < \mathcal{E}' ; \quad E \subset E .$$

An "S'-identity" I' is an (unordered) pair of elements of E':

(2.8) $$I' : \Phi'_1, \Phi'_2 \text{ also written } \Phi'_1 = \Phi'_2 \text{ or } \Phi'_2 = \Phi'_1 .$$

Write

(2.9) $$\mathring{|}(S') = \{\ldots, I', \ldots\} , \quad \mathring{|}(S) = \{\ldots, I, \ldots\}$$

for the class of all S'-(respectively S-) identities. We have

(2.10) $$\mathring{|}(S) \subset \mathring{|}(S') .$$

For a subset $\mathfrak{S}$ of $\mathring{|}(S)$, $\mathfrak{S} \subseteq \mathring{|}(S)$, we define

(2.11) $$|\mathfrak{S}| = \text{"logical closure" of } \mathfrak{S}$$

in the familiar inductive way: (i) $\mathfrak{S} \subseteq |\mathfrak{S}|$. (ii) $\xi_1 = \xi_1 \in |\mathfrak{S}|$. (iii) If $I(\xi_1, \ldots)$: $\Phi(\xi_1, \ldots) = \Psi(\xi_1, \ldots) \in |\mathfrak{S}|$ so also is I^*, where I^* results from I by replacing indeterminate-symbols by S-expressions in any manner. (iv) If $\Phi_1 = \Psi_1$, $\Phi_2 = \Psi_2, \ldots \in |\mathfrak{S}|$ and if $\Pi(\eta, \zeta, \ldots) \in \mathcal{E}$ then $\Pi(\Phi_1, \Phi_2, \ldots) = \Pi(\Psi_1, \Psi_2, \ldots) \in |\mathfrak{S}|$. (v) If $\Phi_1 = \Phi_2$ and $\Phi_2 = \Phi_3 \in |\mathfrak{S}|$ so also is $\Phi_1 = \Phi_3$.

For $\mathfrak{S}' \subseteq \mathring{|}(S')$, $|\mathfrak{S}'|$ is similarly defined. Clearly

$$|\xi_1 = \xi_2| = \mathring{|}\,^{1)} .$$

[1]) Properly we should write $\mathring{|}(S')$, and $\mathring{|}(S) = |\,\xi_1 = \xi_2|$ when only S-operations are admitted. We shall however on occasion write simply $\mathring{|}$ for either of these, where no confusion is possible. A similar remark applies to $\mathring{0}$, $\mathring{0}(S')$, $\mathring{0}(S)$ below.

We define

$$|\xi_1 = \xi_1| = \text{def} = \mathring{0} .$$

Because of (ii), for any $\mathfrak{S}' \subseteq \mathring{|}(S')$,

(2.12) $$\mathring{0} \subseteq |\mathfrak{S}'| \subseteq \mathring{|} .$$

Otherwise said, if $\mathfrak{S}'$ is logically closed,

$$\mathring{0} \subseteq \mathfrak{S}' \subseteq \mathring{|} .$$

By $|\mathfrak{S}'|_S$ we mean the subset of all S-identities in $|\mathfrak{S}'|$,

(2.13) $$|\mathfrak{S}'|_S = |\mathfrak{S}'| \cap \mathring{|}(S) .$$

An "$\mathfrak{S}'$-algebra", $\mathfrak{A}'$, is an S'-algebra such that

(2.14) $$|\mathfrak{A}'| \supseteq |\mathfrak{S}'| .$$

If

(2.15) $$|\mathfrak{A}'| = |\mathfrak{S}'|$$

we speak of an "exact" $\mathfrak{S}'$-algebra.

Let $\varkappa$ be a given cardinal and $\xi_1, \xi_2, \ldots, \xi_\varkappa$ a class of indeterminate-symbols of cardinality $\varkappa$. We write $E_\varkappa = E_\varkappa(S)$ for the totality of S-expressions $\Phi(\xi_1, \ldots, \xi_\varkappa)$ in $\xi_1, \ldots, \xi_\varkappa$ and

(2.16) $$\mathscr{E}_\varkappa = \mathscr{E}_\varkappa(S) = (E_\varkappa, \sigma_1, \sigma_2, \ldots)$$

for the corresponding S-algebra, in which $\sigma_1, \sigma_2, \ldots$ are applied formally in $E_\varkappa$. $E'_\varkappa = E_\varkappa(S')$ and

(2.17) $$\mathscr{E}'_\varkappa = (E'_\varkappa, \sigma_1, \sigma_2, \ldots, \sigma'_1, \cdots)$$

are similarly defined. Obviously

(2.18) $$\begin{gathered} E_\varkappa \subset E , \quad \mathscr{E}_\varkappa < \mathscr{E} \\ E'_\varkappa \subset E' , \quad \mathscr{E}'_\varkappa < \mathscr{E}' \\ E_\varkappa \subset E'_\varkappa , \quad \mathscr{E}_\varkappa < \mathscr{E}'_\varkappa . \end{gathered}$$

If $\mathfrak{A}$ is an S-algebra and $\alpha, \ldots$ are specific elements of $\mathfrak{A}$, then if $\Phi(\alpha, \ldots) = \Psi(\alpha, \ldots)$ holds in $\mathfrak{A}$ we speak of a "tabular identity" of $\mathfrak{A}$; here $\Phi(\xi, \ldots)$, $\Psi(\xi, \ldots)$ are of course S-expressions. We write $\|\mathfrak{A}\|$ for the totality of tabular identities of $\mathfrak{A}$.

3. Compatibility

Def. 3.1. Let $\mathfrak{B}$ be an S-algebra and let $\mathfrak{S}' \subseteq \mathring{|}(S')$. We say that $\mathfrak{S}'$ is "compatible" with $\mathfrak{B}$ — in symbols

$$\mathfrak{B} < \mathfrak{S}'$$

if $\mathfrak{B}$ is imbeddable in some $\mathfrak{S}'$-algebra.

Def. 3.2. Let $\mathfrak{S} \subseteq \mathring{|}(S)$ and $\mathfrak{S}' \subseteq \mathring{|}(S')$. We say that $\mathfrak{S}'$ is compatible with $\mathfrak{S}$ — in symbols

$$\mathfrak{S} \prec \mathfrak{S}'$$

if there exists an $\mathfrak{S}'$-algebra $\mathfrak{B}'$ and an exact[2]) $\mathfrak{S}$-algebra $\mathfrak{B}$ such that $\mathfrak{B} < \mathfrak{B}'$.

Def. 3.3. If $\mathfrak{S} \prec \mathfrak{S}'$ and if there exists an exact $\mathfrak{S}$-algebra, $\mathfrak{B}$, which is convertible into an $\mathfrak{S}'$-algebra, $\mathfrak{B}'$ — $\mathfrak{B} \subset\!< \mathfrak{B}'$ — we say that $\mathfrak{S}'$ is "convertibly compatible" with $\mathfrak{S}$, — in symbols

$$\mathfrak{S} \subset\!\prec \mathfrak{S}'.$$

For $\mathfrak{B} \prec \mathfrak{S}'$ respectively for $\mathfrak{S} \prec \mathfrak{S}'$ it is clearly necessary that

$$|\mathfrak{S}'|_S \subseteq |\mathfrak{B}| \quad \text{resp.} \quad |\mathfrak{S}'|_S \subseteq |\mathfrak{S}|. \tag{3.1}$$

In general, however, neither case of (3.1) is sufficient to insure compatibility, as will be shown in § 7. For the limiting case of equality, however, we note

Lemma 3.1. If $|\mathfrak{S}| = |\mathfrak{S}'|_S$ then $\mathfrak{S} \subset\!\prec \mathfrak{S}'$.

Proof. Choose $\mathfrak{B}'$ as an exact $\mathfrak{S}'$-algebra and define $\mathfrak{B}$ as $\mathfrak{B}'$ *qua* S-algebra, from which the lemma follows. We may readily improve the lemma to the following result in terms of the union $\mathfrak{S} \cup \mathfrak{S}'$ of the classes $\mathfrak{S}$ and $\mathfrak{S}'$:

Theorem 3.2. A necessary and sufficient condition for $\mathfrak{S} \subset\!\prec \mathfrak{S}'$ to hold is that $\mathfrak{S} \prec \mathfrak{S} \cup \mathfrak{S}'$.

Proof: Necessity. If $\mathfrak{S} \subset\!\prec \mathfrak{S}'$ then there exists a $\mathfrak{B} \subset\!< \mathfrak{B}'$ where $|\mathfrak{B}| = |\mathfrak{S}|$ and $|\mathfrak{B}'| \supseteq |\mathfrak{S}'|$. Since we are dealing with a conversion it further follows that $|\mathfrak{B}'| \supseteq |\mathfrak{B}| = |\mathfrak{S}|$. Hence $|\mathfrak{B}'| \supseteq |\mathfrak{S} \cup \mathfrak{S}'|$, whence $\mathfrak{S} \prec \mathfrak{S} \cup \mathfrak{S}'$ and the necessity is established. Conversely, suppose $\mathfrak{S} \prec \mathfrak{S} \cup \mathfrak{S}'$. Then by (3.1), $|\mathfrak{S}' \cup \mathfrak{S}|_S \subseteq |\mathfrak{S}|$. But obviously also $|\mathfrak{S}' \cup \mathfrak{S}|_S \supseteq |\mathfrak{S}|$. Hence $|\mathfrak{S}| = |\mathfrak{S}' \cup \mathfrak{S}|_S$ and therefore $\mathfrak{S} \subset\!\prec \mathfrak{S}' \cup \mathfrak{S}$, by lemma 3.1; from this, *a-fortiori*, $\mathfrak{S} \subset\!\prec \mathfrak{S}'$ and the theorem is established.

It is remarked that no result corresponding to lemma 3.1 is available for $\mathfrak{B}, \mathfrak{S}'$; indeed simple examples exist in which $|\mathfrak{B}| = |\mathfrak{S}'|_S$ and where $\mathfrak{S}'$ is not even compatible with $\mathfrak{B}$. In the converse direction, however, one immediately has:

$$\mathfrak{B} \subset\!< \mathfrak{B}' \Rightarrow |\mathfrak{B}| = |\mathfrak{S}'|_S = |\mathfrak{B}'|_S \tag{3.2}$$

where $\mathfrak{S}'$ is any basis for the identities $|\mathfrak{B}'|$.

4. Free $\mathfrak{S}'$-extensions

With each S-algebra $\mathfrak{B}$ and each class $\mathfrak{S}'(\mathfrak{S}' \subseteq \mathring{|}(S'))$ of S'-identities we associate a certain S'-algebra, $\mathfrak{B}[\mathfrak{S}']$, called the "free $\mathfrak{S}'$-algebra over $\mathfrak{B}$". This algebra will serve as an effective concept in the study and characterization of compatibility.

Consider first a brief intuitive account of the algebra $\mathfrak{B}[\mathfrak{S}']$. The elements of $\mathfrak{B}[\mathfrak{S}']$ are the formal S'-expressions $\Phi'(\beta, \ldots)$ in the elements $\beta, \ldots$ of $\mathfrak{B}$,

[2]) If the restriction "exact" were deleted we should (trivially) have $\mathfrak{S} \prec \mathfrak{S}'$ for all $\mathfrak{S}$ and $\mathfrak{S}'$, since $\mathring{|}(S) \subset\!\prec \mathring{|}(S')$, where $\mathring{|}$ denotes the one-element algebra.

where two such formal expressions are regarded as the same element of $\mathfrak{B}[\mathfrak{S}']$ — in symbols

$$\Phi'(\beta, \ldots) = \Psi'(\beta, \ldots) \qquad (\mathfrak{B}[\mathfrak{S}']) \tag{4.1}$$

if the one is a "$\mathfrak{B}, \mathfrak{S}'$ consequence" of the other, i.e., a consequence of the assumptions that: (1) $|\mathfrak{S}'|$ hold among all S'-expressions $\Gamma'(\beta, \ldots)$ and (2) $\|\mathfrak{B}\|$ hold among all S-expressions $\Pi(\beta, \ldots)$. The primitive operations of $\mathfrak{B}[\mathfrak{S}']$ are defined to be those of species S', namely $\sigma_1, \sigma_2, \ldots, \sigma'_1, \ldots$, applied formally to the $\Phi'(\beta, \ldots)$.

We turn to a precise definition of $\mathfrak{B}[\mathfrak{S}']$. Let $\mathfrak{B}$ be of cardinality $\varkappa$, $\mathfrak{B} = \{\beta_1, \beta_2, \ldots, \beta_\varkappa\}$[3]). It is here convenient to use these symbols β_i — instead of $\xi_1, \ldots, \xi_\varkappa$ as heretofore — as the $\varkappa$ generators of $E'_\varkappa = (E'_\varkappa, \sigma_1, \sigma_2, \ldots, \sigma'_1, \ldots)$, the free S'-algebra (with no identities)[4]) of $\varkappa$ generators. The elements of $E'_\varkappa$ are then the formal S'-expressions $\Phi'(\beta_1, \ldots, \beta_\varkappa)$, where formally different such expressions represent different elements of $E'_\varkappa$; the $\sigma_1, \sigma_2, \ldots, \sigma'_1, \ldots$ are applied formally to the elements of $E'_\varkappa$. As before we have

$$\mathscr{E}_\varkappa < \mathscr{E}'_\varkappa; \quad E_\varkappa \subset E'_\varkappa, \tag{4.2}$$

where $\mathscr{E}_\varkappa$ is isomorphically identified as the S-modul of all S-expressions $\Phi(\beta_1, \ldots, \beta_\varkappa)$ of $\mathscr{E}'_\varkappa$.

In the algebra $\mathscr{E}'_\varkappa$ we define two ideals (= congruence relations) $\mathscr{I}'_1, \mathscr{I}'_2$ as follows:

$$\Phi'(\beta_1, \ldots, \beta_\varkappa) \equiv \Psi'(\beta_1, \ldots, \beta_\varkappa) \quad (\mathscr{I}'_1) \tag{4.3}$$

if and only if

$$\Phi'(\xi_1, \ldots, \xi_\varkappa) = \Psi'(\xi_1, \ldots, \xi_\varkappa) \in |\mathfrak{S}'| \,. \tag{4.3.1}$$

And

$$\Phi'(\beta_1, \ldots, \beta_\varkappa) \equiv \Psi'(\beta_1, \ldots, \beta_\varkappa) \quad (\mathscr{I}'_2) \tag{4.4}$$

if and only if there exists an S'-expression $\Omega'(\xi, \eta, \ldots)$ and S-expressions $\Phi_1(\beta_1, \ldots, \beta_\varkappa)$, $\Psi_1(\beta_1, \ldots, \beta_\varkappa)$, Φ_2, $\Psi_2, \ldots$ such that

$$\begin{aligned} &\Phi'(\beta_1, \ldots) = (\text{formally}) = \Omega'(\Phi_1(\beta_1, \ldots), \Phi_2(\beta_1, \ldots), \ldots) \\ &\Psi'(\beta_1, \ldots) = (\text{formally}) = \Omega'(\Psi_1(\beta_1, \ldots), \Psi_2(\beta_1, \ldots), \ldots) \\ &\Phi_i(\beta_1, \ldots) = \Psi_i(\beta_1, \ldots) \in \|\mathfrak{B}\| \,, \quad (i = 1, 2, \ldots) \,. \end{aligned} \tag{4.4.1}$$

That $\mathscr{I}'_1$ is an ideal in $\mathscr{E}'_\varkappa$ follows at once from the definition of logical closure, $|\mathfrak{S}'|$. Consider then $\mathscr{I}'_2$. We shall first argue that $\equiv (\mathscr{I}'_2)$ is an equivalence relation in $\mathscr{E}'_\varkappa$, and then further show that it is an ideal.

The reflexivity and symmetry of $\equiv (\mathscr{I}'_2)$ are obvious from the definition of $\equiv (\mathscr{I}'_2)$. The proof of transitivity, however, requires some preparation.

---

[3]) Despite the notation, no restriction on the cardinality $\varkappa$ is intended. It will be seen, in the construction, that $\varkappa$ and $\beta_1, \ldots, \beta_\varkappa$ could as well be taken to refer to a set of generators of $\mathfrak{B}$, and the $\Phi'(\beta_1, \ldots)$ to be S'-expressions in these generators.

[4]) More precisely, $\mathscr{E}'_\varkappa$ is the $\mathring{0}(S')$-algebra of $\varkappa$-generators.

Let $\Omega'(\xi, \eta, \ldots)$ be an S'-expression. An S'-expression $\Omega'_0(\zeta_1, \zeta_2, \ldots)$ will be called an "S-reduction" of $\Omega'(\xi, \eta, \ldots)$ if (1°) there exist S-expressions $\Pi_i(\xi, \eta, \ldots)$ such that

$$\Omega'(\xi, \eta, \ldots) = \text{formally} = \Omega'_0(\Pi_1(\xi, \eta, \ldots), \Pi_2(\xi, \eta, \ldots), \ldots)$$

and where (2°) each ζ_i occurs precisely once in the S'-expression $\Omega'_0(\zeta_1, \ldots)$.

An S-reduction $\Omega'_0(\zeta_1, \ldots)$ of $\Omega'(\xi, \eta, \ldots)$ is called "S-irreducible" if it possesses (essentially) no further S-reduction, i.e., if corresponding to each S-reduction $\Omega'_{00}(\mu_1, \mu_2, \ldots)$ of $\Omega'_0(\zeta_1, \zeta_2, \ldots)$ there exists some arrangement $\zeta_{i_1}, \zeta_{i_2}, \ldots$ of $\zeta_1, \zeta_2, \ldots$ such that

$$\Omega'_{00}(\zeta_{i_1}, \zeta_{i_2}, \ldots) = \text{formally} = \Omega'_0(\zeta_1, \zeta_2, \ldots)\,.$$

For example, if $S = (2{,}2)$, $S' = (2{,}2{,}1)$ and if $\sigma_1, \sigma_2, \sigma'_1$ are more conveniently written as $+$, $\times$, 1, the S'-expression $(\xi^2\eta + 1)\,\eta$ has $(\zeta_1\zeta_2 + 1)\,\zeta_3$ and $(\zeta_1 + 1)\,\zeta_2$ as S-reductions, where the latter is S-irreducible.

Slight reflection shows that: (A) Each S'-expression possesses an S-irreducible reduction $\Omega'_0(\zeta_1, \ldots)$; moreover the latter is unique (up to the letters ζ_i employed). It then further follows from (4.4) and (4.4.1) that: (B) If $\Phi'(\beta_1, \ldots) \equiv \Psi'(\beta_1, \ldots)$ $(\mathscr{I}'_2)$ then Φ' and Ψ' have the same S-irreducible reduction $\Omega'_0(\zeta_1, \ldots)$, and moreover there exist S-expressions $\Phi_1(\beta_1, \ldots)$, $\Psi_1(\beta_1, \ldots)$, $\Phi_2, \Psi_2, \ldots$ satisfying (4.4.1) — with Ω' replaced by Ω'_0. From (B) and from (A) (particularly from the unicity of Ω'_0), the transitivity of $\equiv(\mathscr{I}'_0)$ readily follows. Thus $\equiv(\mathscr{I}'_2)$ is an equivalence relation in $\mathscr{E}'_\varkappa$.

To further show that $\mathscr{I}'_2$ is an ideal in $\mathscr{E}'_\varkappa$, let $\sigma'(\xi, \ldots)$ be any primitive operation of $\mathscr{E}'_\varkappa$, and let

$$\Phi'_{(j)}(\beta_1, \ldots) \equiv \Psi'_{(j)}(\beta_1, \ldots)\ (\mathscr{I}'_2);\ j = 1, 2, \ldots\,. \tag{4.5}$$

We must show that

$$\sigma'(\Phi'_{(1)}, \Phi'_{(2)}, \ldots) \equiv \sigma'(\Psi'_{(1)}, \Psi'_{(2)}, \ldots)\ (\mathscr{I}'_2)\,. \tag{4.6}$$

There exist S'-expressions

$$\Omega'_{(1)}(\xi, \ldots),\ \Omega'_{(2)}(\eta, \ldots), \ldots$$

and S-expressions

$$\Phi_{(j)1}(\beta_1, \ldots, \beta_\varkappa)\,,\quad \Psi_{(j)1}\,,\quad \Phi_{(j)2}\,,\quad \Psi_{(j)2}, \ldots \qquad (j = 1,2,\ldots)$$

such that

$$\begin{aligned} \Phi'_{(j)}(\beta_1, \ldots) &= (\text{formally}) = \Omega'_{(j)}(\Phi_{(j)1}, \Phi_{(j)2}, \ldots) \\ \Psi'_{(j)}(\beta_1, \ldots) &= (\text{formally}) = \Omega'_{(j)}(\Psi_{(j)1}, \Psi_{(j)2}, \ldots) \end{aligned} \qquad (j = 1,2,\ldots) \tag{4.7}$$

and

$$\Phi_{(j)i}(\beta_1, \ldots) = \Psi_{(j)i}(\beta_1, \ldots) \in \|\mathfrak{B}\|\,, \quad (i = 1, 2, \ldots;\ j = 1, 2, \ldots)\,. \tag{4.8}$$

Let

$$\Omega'(\xi_1, \xi_2, \ldots, \eta_1, \eta_2, \ldots) = \text{def} = \sigma'(\Omega'_{(1)}(\xi_1, \xi_2, \ldots), \Omega'_{(2)}(\eta_1, \eta_2, \ldots), \ldots)\,. \tag{4.9}$$

Then Ω' is an S'-expression and

$$(4.10)\quad \begin{aligned} \sigma'(\Phi'_{(1)}, \Phi'_{(2)}, \ldots) &= (\text{formally}) = \Omega'(\Phi_{(1)1}, \Phi_{(1)2}, \ldots, \Phi_{(2)1}, \Phi_{(2)2}, \ldots, \ldots) \\ \sigma'(\Psi'_{(1)}, \Psi'_{(2)}, \ldots) &= (\text{formally}) = \Omega'(\Psi_{(1)1}, \Psi_{(1)2}, \ldots, \Psi_{(2)1}, \Psi_{(2)2}, \ldots, \ldots). \end{aligned}$$

The desired result (4.6) follows from (4.10) and (4.8), whence $\mathscr{I}'_2$ as well as $\mathscr{I}'_1$ is an ideal in $\mathscr{E}'_\varkappa$.

Consider the least upper bound

$$(4.11)\qquad \mathscr{I}' = (\mathscr{I}'_1, \mathscr{I}'_2) = \text{l.u.b. of } \mathscr{I}'_1, \mathscr{I}'_2 .$$

$\mathscr{I}'$ is again an ideal in $\mathscr{E}'_\varkappa$, and in terms of the corresponding factor algebra we finally have the desired

Def. 4.1.

$$(4.12)\qquad \mathfrak{B}[\mathfrak{S}'] = \text{def} = \mathscr{E}'_\varkappa/\mathscr{I}' .$$

The elements of $\mathfrak{B}[\mathfrak{S}']$ are the residue classes of $\mathscr{E}'_\varkappa \bmod \mathscr{I}'$, and for $\Phi'(\beta, \ldots)$, $\Psi'(\beta, \ldots) \in E'_\varkappa$ the following two assertions are equivalent:

$$(4.13)\qquad \begin{aligned} \Phi' &\equiv \Psi'(\mathfrak{B}[\mathfrak{S}']) \\ \Phi' &\equiv \Psi'(\mathscr{I}') . \end{aligned}$$

Furthermore from the definitions of $\mathscr{I}'_1$ and $\mathscr{I}'_2$ one readily has the

Theorem 4.1. $\Phi' \equiv \Psi'(\mathfrak{B}[\mathfrak{S}'])$ is inductively characterized by the following conditions:

(1°) If $\Phi', \Psi' \in E'_\varkappa$ and if $\Phi' = \Psi' \in |\mathfrak{S}'|$, then

$$\Phi'(\beta_1, \ldots, \beta_\varkappa) \equiv \Psi'(\beta_1, \ldots, \beta_\varkappa) \quad (\mathfrak{B}[\mathfrak{S}']) .$$

(2°) If $\Phi, \Psi \in E_\varkappa$ and if $\Phi = \Psi \in \|\mathfrak{B}\|$, then

$$\Phi(\beta_1, \ldots, \beta_\varkappa) \equiv \Psi(\beta_1, \ldots, \beta_\varkappa) \quad (\mathfrak{B}[\mathfrak{S}']) .$$

(3°) If $\Pi'(\xi, \eta, \ldots) \in E'$ and if $\Phi'_i, \Psi'_i \in E'_\varkappa$ with

$$\Phi'_i(\beta_1, \ldots, \beta_\varkappa) = \Psi'_i(\beta_1, \ldots, \beta_\varkappa)(\mathfrak{B}[\mathfrak{S}]) , \quad (i = 1, 2, \ldots) ,$$

then

$$\Pi'(\Phi'_1(\beta_1, \ldots), \Phi'_2(\beta_1, \ldots), \ldots) \equiv \Pi'(\Psi'_1(\beta_1, \ldots), \Psi'_2, \ldots) \ (\mathfrak{B}[\mathfrak{S}']) .$$

(4°) If $\Phi', \Psi', \Gamma' \in E'_\varkappa$ and if $\Phi' \equiv \Psi'(\mathfrak{B}[\mathfrak{S}'])$ and $\Psi' \equiv \Gamma'(\mathfrak{B}[\mathfrak{S}'])$, then $\Phi' \equiv \Gamma'(\mathfrak{B}[\mathfrak{S}'])$.

From (4.11) together with the definition of the least upper bound of ideals we have still another characterization of the concept $\equiv(\mathfrak{B}[\mathfrak{S}'])$,

Lemma 4.2. $\Phi'(\beta_1, \ldots, \beta_\varkappa) \equiv \Psi'(\beta_1, \ldots, \beta_\varkappa)$ $(\mathfrak{B}[\mathfrak{S}'])$ if and only if there exists a connecting chain

$$(4.14)\qquad \Phi'(\beta_1, \ldots) \equiv \Gamma'_1(\beta_1, \ldots) \equiv \Gamma'_2(\beta_1, \ldots) \equiv \cdots \equiv \Psi'(\beta_1, \ldots)$$

in which each $\equiv$ denotes $\equiv(\mathscr{I}'_1)$ or else $\equiv(\mathscr{I}'_2)$.

Theorem 4.3. For each S-algebra $\mathfrak{B}$ and each $\mathfrak{S}' \subseteq \uparrow(S')$, $\mathfrak{B}[\mathfrak{S}']$ is an $\mathfrak{S}'$-algebra.

Proof. By (4.11) we have a homomorphism

(4.15) $$\mathcal{E}'_\varkappa/\mathcal{I}'_1 \to \mathcal{E}'_\varkappa/\mathcal{I}' .$$

From the definition of the ideal $\mathcal{I}'$ it is clear that

(4.16) $$\mathcal{E}'_\varkappa/\mathcal{I}'_1 \cong \mathrm{fr}_\varkappa(\mathfrak{S}') = \text{free } \mathfrak{S}'\text{-algebra of } \varkappa\text{-generators}.$$

Hence, using (4.12), we have a homomorphism

(4.17) $$\mathrm{fr}_\varkappa(\mathfrak{S}') \to \mathfrak{B}[\mathfrak{S}] .$$

Since $\mathrm{fr}_\varkappa(\mathfrak{S}')$ is an $\mathfrak{S}'$-algebra and since the identities of an algebra are also identities of any homomorph thereof, the theorem follows.

5. $\mathfrak{B}$-conservation, etc.

An "S-element" of $\mathfrak{B}[\mathfrak{S}']$ is an element — i.e., a residue class of $\mathcal{E}'_\varkappa$ mod $\mathcal{I}'$ (see (4.12)) — in which at least one S-expression $\Phi(\beta_1, \ldots, \beta_\varkappa)$ occurs. If $\overline{B}$ denotes the totality of S-elements of $\mathfrak{B}[\mathfrak{S}']$ we have the

Theorem 5.1. $\overline{\mathfrak{B}} = \text{def} = (\overline{B}, \sigma_1, \sigma_2, \ldots)$ is an S-modul of $\mathfrak{B}[\mathfrak{S}']$,

(5.1) $$\overline{\mathfrak{B}} < \mathfrak{B}[\mathfrak{S}'] .$$

Moreover $\overline{\mathfrak{B}}$ is a homomorphic image of $\mathfrak{B}$,

(5.2) $$\mathfrak{B} \to \overline{\mathfrak{B}} .$$

Proof. That $\overline{\mathfrak{B}}$ is an S-modul is obvious. As for the homomorphism, let $\overline{\beta}_i$ be that residue class of $\mathcal{E}'_\varkappa$ mod $\mathcal{I}'$ in which the element ($= S$-expression) β_i occurs, and consider the mapping

(5.3) $$\beta_i \to \overline{\beta}_i \qquad (i = 1, \ldots, \varkappa) .$$

Since for each S-expression $\Phi(\beta_1, \ldots, \beta_\varkappa)$ there exists at least one $\beta_i \in \mathfrak{B}$ such that $\Phi \equiv \beta_i(\mathcal{I}')$ — for instance where $\Phi(\beta_1, \ldots, \beta_\varkappa) = \beta_i \in \|\mathfrak{B}\|$ —, it follows that (5.3) is a mapping of $\mathfrak{B}$ *onto* $\overline{\mathfrak{B}}$,

(5.4) $$\mathfrak{B} \to \overline{\mathfrak{B}} .$$

Moreover this is a homomorphism since it is seen to be a portion of the homomorphism

(5.5) $$\mathcal{E}'_\varkappa/\mathcal{I}'_2 \to \mathfrak{B}[\mathfrak{S}']$$

which in turn follows from (4.11) and (4.12). This completes the proof of theorem 5.1.

Def. 5. If the homomorphism (5.2) is an isomorphism or, what is equivalent, if for all $\beta, \gamma \in \mathfrak{B}$ with $\beta \neq \gamma$ it follows that $\beta \not\equiv \gamma$ ($\mathfrak{B}[\mathfrak{S}']$), we say that $\mathfrak{S}'$ is "$\mathfrak{B}$-conserving".

If $\mathfrak{S}'$ is $\mathfrak{B}$-conserving we shall agree to identify the algebra $\mathfrak{B}$ with its isomorph $\overline{\mathfrak{B}}$; (5.1) then asserts

(5.6) $$\mathfrak{B} < \mathfrak{B}[\mathfrak{S}'] .$$

Theorem 4.3 together with (5.6) establishes the second half ($=$ sufficiency) of the following

Theorem 5.2. Let $\mathfrak{B}$ be an $\mathfrak{S}$-algebra and let $\mathfrak{S}' \subseteq \mathring{\Gamma}(S')$. A necessary and sufficient condition for $\mathfrak{B}$ and $\mathfrak{S}'$ to be compatible is that $\mathfrak{S}'$ be $\mathfrak{B}$-conserving.

Proof. Necessity: Assuming $\mathfrak{B} \prec \mathfrak{S}'$, there exists an S'-algebra $\mathfrak{B}'$ such that

$$\mathfrak{B} < \mathfrak{B}', |\mathfrak{S}'| \subseteq |\mathfrak{B}'| . \tag{5.7}$$

We shall show that $\mathfrak{S}'$ is $\mathfrak{B}$-conserving. By Def. 5, lemma 4.2 and the remark following (5.3) this is equivalent to establishing (A): If α, γ are elements of $\mathfrak{B}$ connected by a chain

$$\alpha \equiv \Gamma'(\beta_1, \ldots, \beta_\varkappa) \equiv \Gamma_2' \equiv \cdots \equiv \gamma \tag{5.8}$$

in which each $\equiv$ denotes $\equiv (\mathscr{I}_1')$ or else $\equiv (\mathscr{I}_2')$, then α and γ denote the same element of $\mathfrak{B}$.

Lemma 5.3. Let $\Gamma'(\beta_1, \ldots, \beta_\varkappa)$ and $\Gamma_*'(\beta_1, \ldots, \beta_\varkappa)$ be S'-expressions such that (i) in $\mathfrak{B}'$, $\Gamma'(\beta_1, \ldots, \beta_\varkappa)$ is an element of $\mathfrak{B}$, say α, and such that (ii) $\Gamma'(\beta_1, \ldots) \equiv \Gamma_*'(\beta_1, \ldots)$ $(\mathscr{I}_1')$ or else $\Gamma' \equiv \Gamma_*' (\mathscr{I}_2')$. Then in $\mathfrak{B}'$, $\Gamma_*'(\beta_1, \ldots) = \alpha$.

Proof. If $\Gamma' \equiv \Gamma_*'(\mathscr{I}_1')$ the lemma follows immediately since $|\mathfrak{B}'| \supseteq |\mathfrak{S}'|$; in fact we then have $\Gamma'(\xi_1, \ldots) = \Gamma_*'(\xi_1, \ldots)$ identically for all $\xi_1, \ldots$ in $\mathfrak{B}'$. If $\Gamma' \equiv \Gamma_*'(\mathscr{I}_2')$ there exists an S'-expression $\Omega'(\xi_1, \xi_2, \ldots)$ and S-expressions $\Gamma_1(\beta_1, \ldots, \beta_\varkappa)$, $\Gamma_{*1}(\beta_1, \ldots, \beta_\varkappa)$, $\Gamma_2(\beta_1, \ldots)$, $\Gamma_{2*}(\beta_1, \ldots)$, ... such that

$$\begin{aligned} \Gamma'(\beta_1, \ldots) &= \text{formally} = \Omega'(\Gamma_1(\beta_1, \ldots), \Gamma_2(\beta, \ldots), \ldots) \\ \Gamma_*'(\beta_1, \ldots) &= \text{formally} = \Omega'(\Gamma_{*1}(\beta_1, \ldots), \Gamma_{*2}, \ldots) \\ \Gamma_j(\beta_1, \ldots) &= \Gamma_{*j}(\beta_1, \ldots) \in \|\mathfrak{B}\| , \quad (j = 1, 2, \ldots) . \end{aligned}$$

By direct substitution the desired conclusion $\Gamma_*'(\beta_1, \ldots) = \alpha$ follows, and the lemma is established.

An obvious inductive repetition of the above lemma to (5.8) establishes (A) and with it theorem 5.2.

If $\mathfrak{B} \prec \mathfrak{S}'$ we speak of $\mathfrak{B} < \mathfrak{B}'$ as a "model" of the former if $|\mathfrak{B}'| \supseteq |\mathfrak{S}'|$. A "closed" model is one in which $\mathfrak{B}'$ is S'-generated by $\mathfrak{B}$, i.e., in which each element of $\mathfrak{B}'$ may be written (in at least one way) as an S'-expression $\Phi'(\beta, \ldots)$ in the $\beta, \ldots \in \mathfrak{B}$. An "exact" model is one in which $|\mathfrak{B}'| = |\mathfrak{S}'|$.

Lemma 5.4. Let $\mathfrak{B} \prec \mathfrak{S}'$, let $\mathfrak{B} < \mathfrak{B}'$ be a model and let $\Gamma'(\beta, \ldots)$, $\Gamma_*'(\beta, \ldots)$ be S'-expressions in the $\beta, \ldots \in \mathfrak{B}$. If $\Gamma' \equiv \Gamma_*'(\mathscr{I}_1')$ or if $\Gamma' \equiv \Gamma_*'(\mathscr{I}_2')$, then $\Gamma(\beta, \ldots) = \Gamma_*'(\beta, \ldots) \in \|\mathfrak{B}'\|$.

The proof of the lemma requires only trivial modification of that of lemma 5.3 and is here omitted. Inductive repetition of lemma 5.4 to (5.8) then yields

Theorem 5.5. Let $\mathfrak{B} \prec \mathfrak{S}'$ and let $\Phi'(\beta, \ldots) \equiv \Psi'(\beta, \ldots)$ $(\mathfrak{B}[\mathfrak{S}'])$. Then in each model $\mathfrak{B} < \mathfrak{B}'$, $\Phi'(\beta, \ldots) = \Psi'(\beta, \ldots) \in \|\mathfrak{B}'\|$.

The terminology "free $\mathfrak{S}'$-algebra over $\mathfrak{B}$" is justified by

Theorem 5.6. Let $\mathfrak{B} \prec \mathfrak{S}'$ and let $\mathfrak{B} < \mathfrak{B}'$ be a closed model thereof. Then $\mathfrak{B}'$ is a homomorphic image of the free $\mathfrak{S}'$-algebra over $\mathfrak{B}$,

$$\mathfrak{B}[\mathfrak{S}'] \to \mathfrak{B}' .$$

Proof. By (4.12), $\mathscr{I}'$ is the kernel of the homomorphism

$$\mathscr{E}_\varkappa' \to \mathfrak{B}[\mathfrak{S}'] ; \quad \mathscr{E}_\varkappa'/\mathscr{I}' \cong \mathfrak{B}[\mathfrak{S}'] .$$

Since we have a closed model it follows that $\mathfrak{B}'$ is a homomorphic image of $\mathrm{fr}_\varkappa(\mathfrak{S}')$ and hence also of $\mathscr{E}'_\varkappa$

$$\mathscr{E}'_\varkappa \to \mathfrak{B}' \; (\mathscr{E}'_\varkappa \to \mathscr{E}'_\varkappa/\mathscr{I}'_1 \cong \mathrm{fr}_\varkappa(\mathfrak{S}') \to \mathfrak{B}') .$$

Let $\mathscr{M}'$ be the corresponding kernel,

$$\mathscr{E}'_\varkappa/\mathscr{M}' \cong \mathfrak{B}' .$$

But by theorem 5.5 the ideal $\mathscr{M}'$ divides $\mathscr{I}'$ ($\mathscr{I}'$ a refinement of $\mathscr{M}'$), from which one has the homomorphism

$$\mathscr{E}'_\varkappa/\mathscr{I}' \to \mathscr{E}'_\varkappa/\mathscr{M}' .$$

This is the desired result $\mathfrak{B}[\mathfrak{S}'] \to \mathfrak{B}'$ and theorem 5.6 is established.

We have the immediate

Corollary. Let $\mathfrak{B} \prec \mathfrak{S}'$ and let $\mathfrak{B} < \mathfrak{B}'$ be a model thereof. Then $\mathfrak{B}'$ possesses a subalgebra which is a homomorphic image of $\mathfrak{B}[\mathfrak{S}']$.

Proof. As such a subalgebra one may take the (closed) subalgebra of $\mathfrak{B}'$ which is S'-generated by $\mathfrak{B}$.

6. Compatibility and $\mathfrak{S}$-conservation, etc.

We turn to the broader question, the compatibility of $\mathfrak{S}$ and $\mathfrak{S}'$ (Def. 3.2). This calls for a slight modification of the considerations governing the imbedding $\mathfrak{B} \prec \mathfrak{S}'$.

Let $\mathfrak{S} \subseteq \lceil(S)$ and $\mathfrak{S}' \subseteq \lceil(S')$ be given. In a manner analogous to that employed in the construction $\mathfrak{B}[\mathfrak{S}']$, we define a certain S'-algebra $\mathfrak{S}[\mathfrak{S}']$, called the "free $\mathfrak{S}'$-algebra over $\mathfrak{S}$". Roughly, the elements of $\mathfrak{S}[\mathfrak{S}']$ are the expressions $\Phi'(\xi, \ldots)$, i.e., the elements of E', where two such expressions denote the same element of $\mathfrak{S}[\mathfrak{S}']$ — in symbols

$$(6.1) \qquad \Phi'(\xi, \ldots) = \Psi'(\xi, \ldots) \quad (\mathfrak{S}[\mathfrak{S}'])$$

if the one is a "restricted $\mathfrak{S}, \mathfrak{S}'$ consequence" of the other, i.e., a consequence of the assumptions that: (1) $|\mathfrak{S}'|$ holds among all $\mathfrak{S}'$-expressions $\Gamma'(\xi, \ldots)$ and (2) $|\mathfrak{S}|$ holds among all S-expressions $\Pi(\xi, \ldots)$. The primitive operations of $\mathfrak{S}[\mathfrak{S}']$ are again the $o_1, o_2, \ldots, o'_1, \ldots$ of the species S', applied formally to the above elements of $\mathfrak{S}[\mathfrak{S}']$.

More specifically, starting with $\mathscr{E}'$ (and $\mathscr{E}$), the free S'-(respectively the free S-) algebra of (arbitrarily many) generators $\xi, \ldots$ and satisfying no identities, and recalling (2.7), $\mathscr{E} < \mathscr{E}'$, we define two ideals (congruence relations) $\mathscr{I}'_1$, $\mathscr{I}'_2$ in $\mathscr{E}'$ as follows:

$$(6.2) \qquad \Phi'(\xi, \ldots) = \Psi'(\xi, \ldots) \quad (\mathscr{I}'_1)$$

if and only if $\Phi'(\xi, \ldots) = \Psi'(\xi, \ldots) \in |\mathfrak{S}'|$; and

$$(6.3) \qquad \Phi'(\xi, \ldots) = \Psi'(\xi, \ldots) \quad (\mathscr{I}'_2)$$

if and only if there exists a $\Omega'(\eta, \zeta, \ldots) \in \mathscr{E}'$ and $\Phi_1(\xi, \ldots), \Psi_1(\xi, \ldots), \Phi_2(\xi, \ldots), \Psi_2(\xi, \ldots), \ldots \in \mathscr{E}$ such that

$$\begin{aligned} \Phi'(\xi, \ldots) &= \text{formally} = \Omega'(\Phi_1(\xi, \ldots), \Phi_2(\xi, \ldots), \ldots) \\ \Psi'(\xi, \ldots) &= \text{formally} = \Omega'(\Psi_1(\xi, \ldots), \Psi_2(\xi, \ldots), \ldots) \\ \Phi_i(\xi, \ldots) &= \Psi_i(\xi, \ldots) \in |\mathfrak{S}| \qquad (i = 1, 2, \ldots) \end{aligned} \tag{6.4}$$

Let $\mathscr{J}'$ be the least upper bound of these ideals,

$$\mathscr{J}' = \text{def} = (\mathscr{J}_1', \mathscr{J}_2') = \text{least upper bound of } \mathscr{J}_1', \mathscr{J}_2' . \tag{6.5}$$

This is again an ideal in $\mathscr{E}'$ and we now define

Def. 6.1. $\qquad \mathfrak{S}[\mathfrak{S}'] = \text{def} = \mathscr{E}'/\mathscr{J}'$.

The elements of $\mathfrak{S}[\mathfrak{S}']$ are then the residue classes of $\mathscr{E}'$ mod $\mathscr{J}'$, and for $\Phi'(\xi, \ldots), \Psi'(\xi, \ldots) \in E'$ the following assertions are equivalent

$$\begin{aligned} \Phi' &= \Psi'(\mathfrak{S}[\mathfrak{S}']) \\ \Phi' &= \Psi'(\mathscr{J}') . \end{aligned} \tag{6.6}$$

Again, exactly as in lemma 4.2, we have the two results

Lemma 6.1. $\Phi'(\xi, \ldots) = \Psi'(\xi, \ldots)$ $(\mathfrak{S}[\mathfrak{S}'])$ if and only if there exists a connecting chain

$$\Phi'(\xi, \ldots) \equiv \Gamma_1'(\xi, \ldots) \equiv \Gamma_2'(\xi, \ldots) \equiv \cdots \equiv \Psi'(\xi, \ldots) \tag{6.7}$$

in which each $\equiv$ denotes $\equiv (\mathscr{J}_1')$ or else $(\mathscr{J}_2')$.

Theorem 6.2. For each $\mathfrak{S} \subseteq \mathring{\int}(S)$ and $\mathfrak{S}' \subseteq \mathring{\int}(S')$; $\mathfrak{S}[\mathfrak{S}']$ is an $\mathfrak{S}'$-algebra.

It is convenient to express (6.6) in still another way. With each pair $\mathfrak{S}, \mathfrak{S}'$ of classes of identities as above is associated the class of their "restricted (logical) consequences" — written $[\mathfrak{S}, \mathfrak{S}']$,

$$[\mathfrak{S}, \mathfrak{S}'] \subseteq \mathring{\int}(S')$$

where the class $[\mathfrak{S}, \mathfrak{S}']$ is again defined by (6.6):

$$\Phi'(\xi, \ldots) = \Psi'(\xi, \ldots) \in [\mathfrak{S}, \mathfrak{S}'] , \qquad \text{equiv. } \Phi' = \Psi'(\mathscr{J}') . \tag{6.8}$$

From the definition of $\mathscr{J}'$ together with the notion of logical closure one has, in particular,

$$\begin{aligned} |\mathfrak{S}| &\subseteq [\mathfrak{S}, \mathfrak{S}'] \\ |\mathfrak{S}'| &\subseteq [\mathfrak{S}, \mathfrak{S}'] \\ [\mathfrak{S}, \mathfrak{S}'] &\subseteq |[\mathfrak{S}, \mathfrak{S}']| \\ |[\mathfrak{S}, \mathfrak{S}']| &= |\mathfrak{S} \cup \mathfrak{S}'| \qquad (\cup = \text{union}) . \end{aligned} \tag{6.9}$$

It is to be noted that the class of restricted logical consequences of $\mathfrak{S}$ and $\mathfrak{S}'$ is generally not logically closed, that is, in general the strict inclusion $\subset$ obtains in the penultimate relation (6.9). Indeed it may well happen that

$$[\mathfrak{S}, \mathfrak{S}'] \subset |[\mathfrak{S}, \mathfrak{S}']| = \mathring{\int}(S') \, !$$

We write $[\mathfrak{S}, \mathfrak{S}']_S$ for the class of all S-identities occurring in $[\mathfrak{S}, \mathfrak{S}']$,

$$[\mathfrak{S}, \mathfrak{S}']_S = \text{def} = [\mathfrak{S}, \mathfrak{S}'] \cap \mathring{\int}(S) . \tag{6.10}$$

By (6.9) we always have

$$(6.11) \qquad |\mathfrak{S}| \subseteq [\mathfrak{S}, \mathfrak{S}']_S$$

or, equivalently: if $\Phi(\xi, \ldots)$ and $\Psi(\xi, \ldots)$ are S-expressions such that $\Phi = \Psi \in |\mathfrak{S}|$ then $\Phi \equiv \Psi(\mathcal{J}')$.

Def. 6.2. If $|\mathfrak{S}| = [\mathfrak{S}, \mathfrak{S}']_S$ — or equivalently, if for all S-expressions $\Phi(\xi, \ldots)$, $\Psi(\xi, \ldots)$ such that $\Phi \equiv \Psi(\mathcal{J}')$ it follows that $\Phi = \Psi \in |\mathfrak{S}|$ — we say that "$\mathfrak{S}'$ is $\mathfrak{S}$-conserving".

Theorem 6.3. Let $\mathfrak{S} \subseteq \mathring{\rceil}(S)$, $\mathfrak{S}' \subseteq \mathring{\rceil}(S')$. A necessary and sufficient condition for $\mathfrak{S} \prec \mathfrak{S}'$ is that $\mathfrak{S}'$ be $\mathfrak{S}$-conserving.

Proof. (a) We note that the elements of $\mathrm{fr}(\mathfrak{S})$, the free $\mathfrak{S}$-algebra with (arbitrarily many) generators, $\xi, \ldots$ consists of all S-expressions $\Phi(\xi, \ldots)$ where two such, Φ and Ψ, represent the same element of $\mathrm{fr}(\mathfrak{S})$ if and only if $\Phi = \Psi \in |\mathfrak{S}|$.

(b) An "S-element" of $\mathcal{E}'/\mathcal{J}'$ is a residue class of $\mathcal{E}'$, mod $\mathcal{J}'$, in which at least one S-expression $\Phi(\xi, \ldots)$ occurs. The S-elements form an S-modul (the "natural" S-modul) of $\mathcal{E}'/\mathcal{J}'$, i. e., an S-modul of $\mathfrak{S}[\mathfrak{S}']$. Moreover if $\mathfrak{S}'$ is $\mathfrak{S}$-conserving this natural S-modul is clearly isomorphic — and will be identified with — $\mathrm{fr}(\mathfrak{S})$.

(c) For a sufficiently large (cardinal) number of generators each free algebra is exact. Hence $\mathrm{fr}(\mathfrak{S})$ is an exact $\mathfrak{S}$-algebra.

The observations (a)—(c) show that: if $\mathfrak{S} \prec \mathfrak{S}'$,

$$(6.12) \qquad \mathrm{fr}(\mathfrak{S}) < \mathfrak{S}[\mathfrak{S}'] .$$

This together with theorem 6.2 establishes the sufficiency part of the theorem.

As for the necessity, let $\mathfrak{S} \prec \mathfrak{S}'$ and let $\mathfrak{B} < \mathfrak{B}'$ be a model. Then $\mathfrak{S}'$ is $\mathfrak{B}$-conserving, by theorem 5.2, and *a-fortiori* $\mathfrak{S}'$ is $\mathfrak{S}$-conserving. This completes the proof of theorem 6.3.

Let $\mathfrak{A}$ be an S-modul of $\mathfrak{A}'$, $\mathfrak{A} < \mathfrak{A}'$. We call $\Phi'(\xi, \ldots) = \Psi'(\xi, \ldots) \in \mathring{\rceil}(S')$ a "restricted identity" of $\mathfrak{A}'$ if this is identically true in $\mathfrak{A}'$ for all $\xi, \ldots \in \mathfrak{A}$.

Let $\mathfrak{S} \prec \mathfrak{S}'$. We call $\mathfrak{B} < \mathfrak{B}'$ a "model" of $\mathfrak{S} \prec \mathfrak{S}'$ if $|\mathfrak{B}| \supseteq |\mathfrak{S}|$, $|\mathfrak{B}'| \supseteq |\mathfrak{S}'|$; if $|\mathfrak{B}| = |\mathfrak{S}|$ we speak of a left-model.

The restricted consequences $[\mathfrak{S}, \mathfrak{S}']$ are related to the above by

Theorem 6.4. Let $\mathfrak{S} \prec \mathfrak{S}'$ and let $\mathfrak{B} < \mathfrak{B}'$ be any model thereof. Then each restricted consequence of $\mathfrak{S}$ and $\mathfrak{S}'$ is a restricted identity in $\mathfrak{B}'$.

The proof is entirely parallel to that of theorem 5.5 and will be omitted.

While the compatibility of $\mathfrak{S}'$ with $\mathfrak{S}$ does not insure the imbeddability of a preassigned $\mathfrak{S}$-algebra in an $\mathfrak{S}'$-algebra, at least for the free $\mathfrak{S}$-algebra we have

Theorem 6.5. Let $\mathfrak{S} \prec \mathfrak{S}'$. Then for each cardinal $\varkappa$, $\mathrm{fr}_\varkappa(\mathfrak{S})$, the free $\mathfrak{S}$-algebra of $\varkappa$ generators, is imbeddable in an $\mathfrak{S}'$-algebra; indeed, uniformly for all $\varkappa$,

$$(6.13) \qquad \mathrm{fr}_\varkappa(\mathfrak{S}) < \mathfrak{S}[\mathfrak{S}'] .$$

Proof. This is a consequence of (6.12) together with the observation that for each $\varkappa$, $\mathrm{fr}_\varkappa(\mathfrak{S})$ is a subalgebra of $\mathrm{fr}(\mathfrak{S})$.

Although $\mathrm{fr}_\varkappa(\mathfrak{S})$ is an exact $\mathfrak{S}$-algebra for $\varkappa$ sufficiently large, this need not be true for small $\varkappa$. Theorem 6.5 may then be restated as a characterization of compatibility:

Theorem 6.5.1. A necessary and sufficient condition for $\mathfrak{S} \prec \mathfrak{S}'$ is that each free $\mathfrak{S}$-algebra be imbeddable in an $\mathfrak{S}'$-algebra.

7. An "insufficiency" example

For $\mathfrak{S} \prec \mathfrak{S}'$ it is necessary that $|\mathfrak{S}'|_S \subseteq |\mathfrak{S}|$. By means of a simple example we shall show that this condition is generally not sufficient.

Take $S = (2)$, $S' = (2,1)$ and for convenience temporarily write $\xi\eta$ (or $\xi \times \eta$) and ξ^* in place of $\sigma_1(\xi, \eta)$ and $\sigma_1'(\xi)$ respectively. As $\mathfrak{S}$ and $\mathfrak{S}'$ take

$$\mathfrak{S}\begin{cases} \xi(\eta\zeta) = (\xi\eta)\zeta \\ \xi^2 = \xi^3 \end{cases} \tag{7.1}$$

$$\mathfrak{S}'\begin{cases} \xi = \xi^{2*} (= (\xi^2)^*) \\ \xi^2 = (\xi^2 \times \xi)^* . \end{cases} \tag{7.2}$$

We shall show that $|\mathfrak{S}'|_S \subset |\mathfrak{S}|$ and further, that $\mathfrak{S}$ and $\mathfrak{S}'$ are not compatible.

Lemma 7. $|\mathfrak{S}'|_S$ is "empty" (i.e., $= \mathring{0}(S)$).

Proof. Let $\varkappa$ be a given cardinal number and let $\mathrm{fr}_\varkappa(\mathring{0})$ be the free $\mathring{0}(S)$-algebra (i.e., no identities) of $\varkappa$-generators $\alpha_1, \ldots, \alpha_\varkappa$. We first show that $\mathrm{fr}_\varkappa(\mathring{0})$ may be *converted* into an $\mathfrak{S}'$-algebra.

The elements of $\mathrm{fr}_\varkappa(\mathring{0})$ are the formal $\times$-expressions $\Phi(\alpha_1, \ldots, \alpha_\varkappa)$ in $\alpha_1, \ldots, \alpha_\varkappa$; in particular the α_i themselves are special $\in \mathrm{fr}_\varkappa(\mathring{0})$. Each element Π different from the α_i is *uniquely* factorable

$$\Pi = \Phi \times \Psi . \tag{7.3}$$

A "squared" element is one factorable in the form $\Phi \times \Phi$; a "semi-cubed" element is one factorable in the form $(\Psi \times \Psi) \times \Psi$. From the unicity of the factorization (7.3) follows: (i) $\Phi_1 \times \Phi_1 = \Phi_2 \times \Phi_2$ implies $\Phi_1 = \Phi_2$, (ii) $(\Psi_1 \times \Psi_1) \times \Psi_1 = (\Psi_2 \times \Psi_2) \times \Psi_2$ implies $\Psi_1 = \Psi_2$, (iii) $\Phi \times \Phi = (\Psi \times \Psi) \times \Psi$ is impossible.

If ξ is a squared element, $\xi = \Phi \times \Phi$, we define $\xi^* = \Phi$; if ξ is a semi-cubed element, $\xi = (\Psi \times \Psi) \times \Psi$, we define $\xi^* = \Psi \times \Psi$; for all other ξ we define $\xi^* = \xi$. In view of (i)—(iii) above, ξ^* is uniquely defined for each $\xi \in \mathrm{fr}_\varkappa(\mathring{0})$; moreover it is immediately verified that $\mathfrak{S}'$ is satisfied. Hence we have a conversion of $\mathrm{fr}_\varkappa(\mathring{0})$ — call it $\mathrm{fr}_\varkappa'$ — into an $\mathfrak{S}'$-algebra,

$$\mathrm{fr}_\varkappa(\mathring{0}) \subset\!< \mathrm{fr}_\varkappa' . \tag{7.4}$$

From this follows

$$|\mathrm{fr}_\varkappa'| \supseteq |\mathfrak{S}'| , \quad |\mathrm{fr}_\varkappa'|_S \supseteq |\mathfrak{S}'|_S . \tag{7.5}$$

Since

$$|\mathrm{fr}_\varkappa'|_S = |\mathrm{fr}_\varkappa(\mathring{0})| = \mathring{0}(S) \tag{7.6}$$

we have $|\mathfrak{S}'|_S = \mathring{0}$, which establishes the lemma.

From the lemma we then have

$$|\mathfrak{S}'|_S \subset |\mathfrak{S}| . \tag{7.7}$$

There remains to show that $\mathfrak{S}'$ and $\mathfrak{S}$ are not compatible. In the algebra $\mathfrak{S}[\mathfrak{S}']$, with $\mathscr{I}_1'$ and $\mathscr{I}_2'$ defined as in (6.2) and (6.3), we have the chain

$$\xi^2 = (\xi^2 \times \xi)^*(\mathscr{I}_1') = \xi^{3*}(\mathscr{I}_2') = \xi^{2*}(\mathscr{I}_2') = \xi(\mathscr{I}_1') . \tag{7.8}$$

That is,

$$\xi^2 = \xi(\mathfrak{S}[\mathfrak{S}']) . \tag{7.9}$$

Since clearly

$$\xi^2 = \xi \notin |\mathfrak{S}| \tag{7.10}$$

it follows that $\mathfrak{S}'$ is not $\mathfrak{S}$-conserving, or equivalently, $\mathfrak{S}'$ and $\mathfrak{S}$ are not compatible. This result together with (7.7) establishes the insufficiency example.

8. "Polynomial extensions"

Within the foregoing framework we touch on the simple class of "polynomial" imbeddings. Let $\mathfrak{B}$ be an S-algebra and let $\mathfrak{S}$ be an identity basis, $|\mathfrak{S}| = |\mathfrak{B}|$. Take

$$S' = (S, 1) = (n_1, n_2, \ldots, n_1'(= 1)) , \tag{8.1}$$

let $\mathfrak{S}_0$ be a fixed subset of $|\mathfrak{S}|$,

$$\mathring{0}(S) \subseteq \mathfrak{S}_0 \subseteq |\mathfrak{S}_0| \subseteq |\mathfrak{S}| = |\mathfrak{B}| \tag{8.2}$$

and let $\mathfrak{S}_0'$ be defined as

$$\mathfrak{S}_0' = \mathfrak{S}_0 \text{ augmented by the constant identity: } \sigma_1'(\xi) = \sigma_1'(\eta) . \tag{8.3}$$

Since $\mathfrak{B}$ is convertible to an $\mathfrak{S}_0'$-algebra in an obvious way, we have

$$\mathfrak{B} < \mathfrak{S}_0' \tag{8.4}$$

from which follows

$$\mathfrak{B} < \mathfrak{B}[\mathfrak{S}_0'] . \tag{8.5}$$

We call $\mathfrak{B}[\mathfrak{S}_0']$ the algebra of all "$\mathfrak{S}_0$-polynomials" over $\mathfrak{B}$ (or with "coefficients" in $\mathfrak{B}$). If in $\mathfrak{B}[\mathfrak{S}_0']$ we write simply χ for the fixed element $\sigma_1'(\xi)$,

$$\sigma_1'(\xi) = \sigma_1'(\eta) = \chi , \tag{8.6}$$

the elements of $\mathfrak{B}[\mathfrak{S}_0']$ are simply the S-expressions $\Phi(\beta, \ldots, \chi)$ in χ and/or the $\beta \in \mathfrak{B}$; we also write simply $\Phi(\chi)$ for these. Two such $\mathfrak{S}_0$-polynomials $\Phi(\chi)$, $\Psi(\chi)$ denote the same element of $\mathfrak{B}[\mathfrak{S}_0']$ only if they are connectable by a chain (4.14), or, more briefly: only if $\Phi(\chi)$ is formally convertable into $\Psi(\chi)$ if one assumes $|\mathfrak{S}_0|$ to hold among all $\mathfrak{S}_0$-polynomials, and also $\|\mathfrak{B}\|$ to hold among all "coefficients" ($= \Phi(\beta, \ldots, \chi)$ in which χ does not occur).

If $\mathfrak{B} = \mathscr{R}$ is taken as a commutative ring with identity, $\mathscr{R} = (R, \times, +, 0, 1)$, $\mathscr{R}[\chi]$, the usual polynomial ring, is recognized as $\mathscr{R}[\mathfrak{S}_0']$ where

$$(8.7)\quad \mathfrak{S}_0 = \begin{cases} \xi + (\eta + \zeta) = (\xi + \eta) + \zeta,\ \xi + \eta = \eta + \xi,\ 0 + \xi = \xi, \\ \qquad \xi(\eta\zeta) = (\xi\eta)\zeta,\ \xi\eta = \eta\xi, \\ 0\zeta = 0,\ 1\xi = \xi,\ \xi(\eta + \zeta) = \xi\eta + \xi\zeta. \end{cases}$$

In the extreme case where

$$(8.8)\qquad |\mathfrak{S}_0| = |\mathfrak{S}|\ (= |\mathfrak{B}|),$$

$\mathfrak{B}[\mathfrak{S}_0']$ is seen to be isomorphic with the algebra of all $\mathfrak{B}$-functions $\Phi(\chi)$: $\Phi(\chi) = \Psi(\chi)\ (\mathfrak{B}[\mathfrak{S}_0'])$ if and only if Φ and Ψ represent the same self-mapping in $\mathfrak{B}$. Indeed in the general case (8.2), if the $\Phi(\chi)$ are interpreted as mappings in $\mathfrak{B}$, half of the above still holds: $\Phi(\chi) = \Psi(\chi)\ (\mathfrak{B}[\mathfrak{S}_0'])$ implies Φ and Ψ denote the same mappings. When $|\mathfrak{S}_0| \neq |\mathfrak{S}|$, however, the converse is generally false. It is apparent that: the algebra of all $\mathfrak{B}$-functions (in one argument) is a homomorphic map of $\mathfrak{B}[\mathfrak{S}_0']$, for each $|\mathfrak{S}_0'| \subseteq |\mathfrak{S}|$.

Finally we remark that the algebra of $\mathfrak{S}_0$-polynomials over $\mathfrak{B}$ in more than one letter are similarly obtained by selecting $S''\cdots\ (S, 1, 1, \ldots)$, and $\sigma_j'(\xi) = \sigma_j'(\eta)\ (j = 1, 2, \ldots)$.

9. Universal compatibility

From the compatibility $\mathfrak{S} \prec \mathfrak{S}'$ we cannot generally conclude that each exact $\mathfrak{S}$-algebra — and *a-fortiori* that each $\mathfrak{S}$-algebra — is imbeddable in an $\mathfrak{S}'$-algebra. We augment the definitions of § 3 with

Def. 9. Let $\mathfrak{S} \subseteq \mathring{\upharpoonright}(S)$ and $\mathfrak{S}' \subseteq \mathring{\upharpoonright}(S')$. If $\mathfrak{B} \prec \mathfrak{S}'$ for each $\mathfrak{S}$-algebra $\mathfrak{B}$ we say that $\mathfrak{S}'$ is "universally compatible" or "linked" with $\mathfrak{S}$, — in symbols,

$$\mathfrak{S} \prec\prec \mathfrak{S}'.$$

From theorem 5.2 we have

$$(9.1)\qquad \mathfrak{S} \prec\prec \mathfrak{S}' \Leftrightarrow \mathfrak{B} \prec \mathfrak{B}[\mathfrak{S}'] \text{ for each } \mathfrak{S}\text{-algebra } \mathfrak{B}.$$

We further note the transitivity: For $S \subset S' \subset S''$ and $\mathfrak{S}'' \subseteq \mathring{\upharpoonright}(S'')$,

$$(9.2)\qquad \mathfrak{S} \prec\prec \mathfrak{S}' \text{ and } \mathfrak{S}' \prec\prec \mathfrak{S}'' \Rightarrow \mathfrak{S} \prec\prec \mathfrak{S}''.$$

The universal imbeddability of a ring in a ring with identity is easily recast as an instance of linkage between suitable $\mathfrak{S}$ and $\mathfrak{S}'$, and a similar remark applies to a number of additional imbedding situations related to rings. On a broader (universal algebra) front, with a minimum of specification of algebraic structure, theorems on linkage are not so easily come by. Theorem 9.1 and, more importantly, sections 10 and 11 bear on this situation.

If in $I' \in \mathring{\upharpoonright}(S')$ precisely one member, Φ, is an S-expression, we speak of an "$S - S'$ identity",

$$(9.3)\qquad I' : \Phi = \Psi'.$$

Let $|\mathfrak{S}'|_{SS'}$ be the subset of all $S - S'$ identities in $|\mathfrak{S}'|$; if this is empty we write $|\mathfrak{S}'|_{SS'} = \Lambda$.

Theorem 9.1. Let $\mathfrak{S} \subseteq \rceil(S)$, $\mathfrak{S}' \subseteq \rceil(S')$, $|\mathfrak{S}'|_{SS'} = \Lambda$ and $\mathfrak{S} < \mathfrak{S}'$. Then $\mathfrak{S} < < \mathfrak{S}'$.

Proof. Let $\mathfrak{B}$ be any $\mathfrak{S}$-algebra, $|\mathfrak{B}| \geqq |\mathfrak{S}|$. We must show that $\mathfrak{B} < \mathfrak{B}[\mathfrak{S}']$ or equivalently: if α and γ are $\in \mathfrak{B}$ which are connected by a chain (5.8), then $\alpha = \gamma$ in $\mathfrak{B}$.

We establish (A): If $\Gamma_i(\beta, \ldots) \equiv \Gamma'_{i+1}(\beta, \ldots)$ where $\equiv$ denotes $\equiv (\mathscr{I}'_1)$ or else $\equiv (\mathscr{I}'_2)$ and where Γ_i is an S-expression in the $\beta, \ldots \in \mathfrak{B}$, then Γ'_{i+1} is also an S-expression and $\Gamma_i = \Gamma'_{i+1} \in \|\mathfrak{B}\|$. For $\equiv (\mathscr{I}'_1)$ we prove (A) as follows. In this case $\Gamma_i = \Gamma'_{i+1} \in |\mathfrak{S}'|$, and from $|\mathfrak{S}'|_{SS'} = \Lambda$ follows that Γ'_{i+1} is also an S-expression, and $\Gamma_i = \Gamma'_{i+1} \in |\mathfrak{S}'|_S$. Further from $\mathfrak{S} < \mathfrak{S}'$ follows $|\mathfrak{S}'|_S \subseteq |\mathfrak{S}|$, whence $\Gamma_i = \Gamma'_{i+1} \in |\mathfrak{S}|$ and *a-fortiori* $\Gamma_i = \Gamma'_{i+1} \in \|\mathfrak{B}\|$. Again, for $\equiv (\mathscr{I}'_2)$ we have

$$(9.4)\quad \begin{aligned} \Gamma_i(\beta, \ldots) &= \text{formally} = \Omega'(\Gamma_{i1}(\beta, \ldots), \Gamma_{i2}(\beta, \ldots), \ldots) \\ \Gamma'_{i+1}(\beta, \ldots) &= \text{formally} = \Omega'(\Gamma_{i+1,1}(\beta, \ldots), \Gamma_{i+1,2}, \ldots), \end{aligned}$$

where $\Omega'(\xi, \eta, \ldots)$ is an S'-expression and $\Gamma_{i1}, \ldots, \Gamma_{i+1,1}, \ldots$ are S-expressions in the $\beta, \ldots \in \mathfrak{B}$, with $\Gamma_{ij} = \Gamma_{i+1,j} \in \|\mathfrak{B}\|$ $(j = 1, 2, \ldots)$. From the first of equations (9.4) it follows that Ω' is an S-expression, and thus again $\Gamma_i = \Gamma'_{i+1} \in \|\mathfrak{B}\|$. This establishes (A).

Since α is itself an S-expression in the $\in \mathfrak{B}$, successive application of (A) to the chain (5.8) shows that $\alpha = \gamma$ in $\mathfrak{B}$. This completes the proof of theorem 9.1. Along similar lines — we omit the details — one may throw theorem 9.1 into the form

Theorem 9.2. Let $\mathfrak{S} \subseteq \rceil(S)$, $\mathfrak{S}' \subseteq \rceil(S')$, $|\mathfrak{S}'|_{SS'} = \Lambda$ and $|\mathfrak{S}'|_S \subseteq |\mathfrak{S}|$. Then $\mathfrak{S} < < \mathfrak{S}'$.

A simple instance is furnished if $\mathfrak{S}'$ is a class of S'-S' identities $\Phi'_i = \Psi'_i$, i.e., where neither Φ'_i nor Ψ'_i is an S-expression; the hypotheses of the theorem are then obviously satisfied. We turn to two more interesting cases of linkage.

10. Primal sets and universal compatibility

We call $\mathfrak{S} \subset \rceil(S)$ a "primal set" of identities, of species S, if for some primal algebra $\mathfrak{P}$ we have that

$$|\mathfrak{S}| = |\mathfrak{P}|.$$

Otherwise stated, $\mathfrak{S}$ is a primal set if and only if it forms a basis for the identities of some primal algebra, $\mathfrak{P}$. We recall (see [2], [3]) that a primal algebra is a finite algebra which is functionally strictly complete, and not isomorphic with the one-element algebra. From a fundamental theorem of [2], [3] it follows that such an associated primal $\mathfrak{P}$ is unique: to each primal set $\mathfrak{S}$ corresponds precisely one primal S-algebra, $\mathfrak{P}$.

Primal algebras — and hence primal sets — occur in profusion. They embrace, for example, all n-fields (which in turn, for $n = p^k$ = prime power, include a particular representation of all Galois fields), all basic Post-algebras $\mathscr{P}_n$, all finite groups with null-element, etc. — compare with [1]. In particular,

for each species S (except $S = (1)$) and for each integer $n > 1$ there exists at least one primal S-algebra of order n.

Theorem 10.1. Let $\mathfrak{S} \subset \mathring{|}(S)$, $\mathfrak{S}' \subset \mathring{|}(S')$, $\mathfrak{S} \prec \mathfrak{S}'$ and let $\mathfrak{S}$ be a primal set. Then $\mathfrak{S}$ and $\mathfrak{S}'$ are linked, $\mathfrak{S} \prec\prec \mathfrak{S}'$.

Proof. Let $\mathfrak{P}$ be the (unique) primal algebra belonging to $\mathfrak{S}$. Since $\mathfrak{S} \prec \mathfrak{S}'$ we have $\mathfrak{B} < \mathfrak{B}'$ for some $\mathfrak{S}'$-algebra $\mathfrak{B}'$ and some exact $\mathfrak{S}$-algebra $\mathfrak{B}$. We now appeal to the first fundamental structure theorem of primal algebras (see [3]) which asserts (in part): Each $\mathfrak{S}$-algebra is isomorphic with some subdirect power of $\mathfrak{P}$; moreover since $\mathfrak{B}$ above is an exact $\mathfrak{S}$-algebra, $\mathfrak{B} \neq \mathring{|}(S)$ (= one-element S-algebra = 0^{th} direct power of $\mathfrak{P}$) It is a further result of [2] that: every subdirect power (except the 0^{th}) of a primal algebra $\mathfrak{P}$ is "normal" and therefore each subdirect power (except the 0^{th}) contains a subalgebra isomorphic with $\mathfrak{P}$. Thus it follows that

$$\mathfrak{P} < \mathfrak{B}' \tag{10.1}$$

From (10.1) we thus have that: every subalgebra of a direct power of $\mathfrak{P}$ (and hence in particular every subdirect power of $\mathfrak{P}$) is imbeddable in the corresponding direct power of $\mathfrak{B}'$. By a second appeal to the fundamental structure theorem we have that: every $\mathfrak{S}$-algebra is imbeddable in a direct power of $\mathfrak{B}'$. Since a direct power of $\mathfrak{B}'$ is necessarily an $\mathfrak{S}'$-algebra and since the exceptional $\mathfrak{S}$-algebra $\mathring{|}(S)$ is obviously imbeddable in (convertible to) the $\mathfrak{S}'$-algebra $\mathring{|}(S')$, the theorem is proved. The above proof actually establishes the stronger

Theorem 10.2. Let $\mathfrak{S} \subset \mathring{|}(S)$, $\mathfrak{S}' \subset \mathring{|}(S')$, $\mathfrak{S} \prec \mathfrak{S}'$ and let $\mathfrak{S}$ be a primal set of identities. Then there exists an $\mathfrak{S}'$-algebra, $\mathfrak{B}'$, such that each $\mathfrak{S}$-algebra is imbeddable in some direct power of $\mathfrak{B}'$.

If in theorem 10.2 $\mathfrak{S}'$ is also primal — with the corresponding primal algebra denoted by $\mathfrak{Q}'$ — then, by a further application of the structure theorem, $\mathfrak{B}'$ is isomorphic with a subdirect power of $\mathfrak{Q}'$. Here the most interesting subcase is where $\mathfrak{P} < \mathfrak{Q}'$.

11. Categorical algebras and universal compatibility

A slight extension of the notion of "$\mathfrak{S}$-algebra" is here convenient. If $\mathfrak{B}$ is an S-algebra, by a "$\mathfrak{B}$-algebra" we mean simply a $|\mathfrak{B}|$-algebra, i.e., an S-algebra which satisfies all identities $|\mathfrak{B}|$ of $\mathfrak{B}$.

Def. 11.1. An S-algebra, $\mathfrak{A}$, is said to be "categorical" if it is a finite algebra not isomorphic with the one-element algebra $\mathring{|}(S)$, and if each $\mathfrak{A}$-algebra is isomorphic with a subdirect power of $\mathfrak{A}$.

From the fundamental structure theorem referred to in § 10, a primal algebra is seen to be categorical. Indeed it will presently be seen that various properties of primal algebras are shared by categorical algebras. That this latter class properly subsumes the class of primals is clear from the simple

example of $(C_p, \overset{t}{+})$, the cyclic group of prime order p, which is categorical but not primal — (see remark 2 following theorem 11.4).

In this section we shall exhibit a characterization of categorical algebras in terms of a linkage (universal compatibility) of their identities with the identities of imbedding primal algebras. We require some preparation.

We shall say that an algebra $\mathfrak{A}$ possesses "essentially no subalgebra" if $\mathfrak{A}$ either possesses no subalgebra (different from $\mathfrak{A}$ itself) or else if $\mathfrak{A}$ possesses precisely one subalgebra, and where this is a one-element subalgebra (which is then called *the* fixed point of $\mathfrak{A}$). One has the evident

Lemma 11.1. If $\mathfrak{A}$ is an algebra with essentially no subalgebra, then each subalgebra of a direct power of $\mathfrak{A}$ is isomorphic[5]) with a subdirect power of $\mathfrak{A}$.

A "simple" algebra $\mathfrak{A}$ is one possessing no proper ideal, or, equivalently, no proper homomorphism. An S-algebra $\mathfrak{A}$ is "equationally maximal" if $\mathfrak{A} \neq \dot{|}(S)$ and, given $I \in \mathring{|}(S)$, either $I \in |\mathfrak{A}|$ or else $|\,|\mathfrak{A}| \cup I| = \mathring{|}(S)$; equivalently: for $I \notin |\mathfrak{A}|$, the sole S-algebra satisfying $|\mathfrak{A}|$ and also I is $\dot{|}(S)$, the one-element S-algebra (where of course $|\,\dot{|}(S)| = \mathring{|}(S)$).

Theorem 11.2. Let $\mathfrak{A}$ be a categorical S-algebra, where $S(n_1, n_2, \ldots)$ has at least one $n_i \neq 1$. Then (i) $\mathfrak{A}$ is simple, (ii) $\mathfrak{A}$ possesses essentially no subalgebra, (iii) $\mathfrak{A}$ is equationally maximal and (iv) a subalgebra of a direct power of $\mathfrak{A}$ is isomorphic with a subdirect power of $\mathfrak{A}$.

Proof. Clearly (A): A subdirect power of $\mathfrak{A}$ (other than $\mathfrak{A}^{(0)} = \dot{|}$) has no fewer elements than $\mathfrak{A}$. Hence if $\mathfrak{A}_0$ were a proper homomorph of $\mathfrak{A}$ we should have a contradiction with (A), since each homomorph of $\mathfrak{A}$ is a $\mathfrak{A}$-algebra. This proves (i).

Suppose $\mathfrak{A}$ to possess a subalgebra $\mathfrak{A}_1$ different from $\mathfrak{A}$ and not a one-element subalgebra. $\mathfrak{A}_1$ is again a $\mathfrak{A}$-algebra. Moreover since $\mathfrak{A}$ contains a non-unary primitive operation it is readily seen that $\mathfrak{A}_1 \neq \dot{|}(S)$. So the existence of $\mathfrak{A}_1$ violates (A). Since the union of all one-element subalgebras is again a subalgebra, $\mathfrak{A}$ possesses at most one fixed point and (ii) is proved.

The algebra $\mathfrak{A}^{(0)} = \dot{|}$ satisfies all possible S-identities; any other $\mathfrak{A}$-algebra — as subdirect power of $\mathfrak{A}$ — satisfies precisely $|\mathfrak{A}|$. This proves (iii). Finally (iv) is a consequence of (ii) and lemma 11.1, which completes theorem 11.2

Recalling the notion of conversion, (2.3), we require the

Lemma 11.3. Let $\mathfrak{A}$ be a categorical S-algebra and let $\mathfrak{A}'$ be any S'-conversion of $\mathfrak{A}$, $\mathfrak{A} \subset\!< \mathfrak{A}'$. Then each $\mathfrak{A}$-algebra is imbeddable in a $\mathfrak{A}'$-algebra.

Proof. Each $\mathfrak{A}$-algebra is isomorphic with a subdirect power, $\mathfrak{A}^*$, of $\mathfrak{A}$. If $\mathfrak{A}^{**}$ denotes the full direct power of $\mathfrak{A}$ corresponding to $\mathfrak{A}^*$ and if we convert $\mathfrak{A}^{**}$ to $\mathfrak{A}'^{**}$, then clearly $\mathfrak{A} < \mathfrak{A}'^{**}$ and the lemma follows. (It is remarked that, under the given assumptions, a $\mathfrak{A}$-algebra is not generally convertible into a $\mathfrak{A}'$-algebra.)

---

[5]) One cannot generally conclude that each subalgebra of a direct power of $\mathfrak{A}$ *is* necessarily a subdirect power of $\mathfrak{A}$, merely that it is isomorphic with such.

If $\mathfrak{U} \subset< \mathfrak{U}'$ and $\mathfrak{U}'$ is a primal algebra (species S'), we speak of a "primal conversion" (or primal S'-conversion) of $\mathfrak{U}$. If S' contains at least one non-unary primitive operation, then each finite S-algebra ($\neq \dot{|}(S)$) possesses at least one primal S'-conversion.

We combine a portion of lemma 11.3 with a partial converse thereof to obtain the following characterization of the concept of categoricity:

Theorem 11.4. Let $\mathfrak{U}(\mathfrak{U} \neq \dot{|})$ be a finite algebra. Then $\mathfrak{U}$ is categorical if and only if (i) $\mathfrak{U}$ possesses essentially no subalgebra and (ii) there exists a primal conversion $\mathfrak{U}'$ of $\mathfrak{U}$ which is linked to $\mathfrak{U}$

$$|\mathfrak{U}| < < |\mathfrak{U}'|,$$

that is, such that each $\mathfrak{U}$-algebra is imbeddable in a $\mathfrak{U}'$-algebra.

Proof. If $\mathfrak{U}$ is categorical, (i) and (ii) follow from theorem 11.2 and lemma 11.3. As for the sufficiency, let $\mathfrak{U}$ satisfy (i) and (ii) and let $\mathfrak{U}^*$ be a $\mathfrak{U}$-algebra. Since a $\mathfrak{U}'$-algebra is now necessarily isomorphic with a subdirect power of $\mathfrak{U}'$ (by the fundamental structure theorem), we have: $\mathfrak{U}^*$ is imbeddable in a subdirect power of $\mathfrak{U}'$. Since $\mathfrak{U}'$ is a conversion of $\mathfrak{U}$ this is equivalent to: $\mathfrak{U}^*$ is a subalgebra of a subdirect power of $\mathfrak{U}$. Again, since $\mathfrak{U}$ further satisfies (ii), we have from lemma 11.1 that: $\mathfrak{U}^*$ is isomorphic with a subdirect power of $\mathfrak{U}$, i.e., $\mathfrak{U}$ is categorical. This establishes the theorem.

Remark 1. From the limited role of lemma 11.3 in the above proof it is seen that: if there exists a primal conversion $\mathfrak{U}'$ of $\mathfrak{U}$ such that $\mathfrak{U}'$ is linked to $\mathfrak{U}$, then *any* primal conversion of $\mathfrak{U}$ is linked to $\mathfrak{U}$.

Remark 2. The categoricity of $C_p = (C_p, +)$, the cyclic group of prime order, may be established by first converting it to the primal algebra (field) $C'_p = (C_p, +, \times, 1)$ and then easily establishing the linkage $|C_p| < < |C'_p|$. This, together with theorem 11.2, yields a short proof of the familiar equational maximality of $|C|_p$.

Def. 11.2. A finite S-algebra $\mathfrak{B}(\mathfrak{B} \neq \mathring{|})$ is called "semi-categorical" if each $\mathfrak{B}$-algebra is isomorphic with a subalgebra of a direct power of $\mathfrak{B}$.

The class of semi-categorical algebras embraces, in particular, all (i) primal algebras, (ii) categorical algebras, (iii) finite Abelian groups, (iv) $GF(p^t)$ and more generally $GF(p_1^{t_1}) \times GF(p_2^{t_2}) \times \cdots$, the direct product of Galois fields[6]).

Minor modifications of the argument for theorem 11.4 shows that theorem 11.4, with (i) deleted, characterizes semi-categoricity, i.e.,

[6]) Here a precautionary word is in order. Let p, q be distinct primes, let $\mathcal{F}'_p = (F_p, +, \times, 1) = GF(p)$, of species $S' = (2, 2, 1)$, and let $\mathcal{F}_p = (F_p, +, \times) = \mathcal{F}'_p$ *qua* $(+, \times)$-algebra. Then as asserted, $\mathcal{R} = \mathcal{F}_p \times \mathcal{F}_p$ is semi-categorical. However $\mathcal{R}' = \mathcal{F}'_p \times \mathcal{F}'_p$ is not semi-categorical. This stems from the observation that $\mathcal{F}'_p$, as homomorph of $\mathcal{R}'$, is an $\mathcal{R}'$-algebra, whereas $\mathcal{R}'$ possesses no subalgebra and *a-fortiori* no subalgebra isomorphic with $\mathcal{F}'_p$.

Theorem 11.5. A finite algebra $\mathfrak{B}$ is semi-categorical if and only if there exists a primal conversion $\mathfrak{B}'$ of $\mathfrak{B}$ which is linked to $\mathfrak{B}$, i.e., such that $|\mathfrak{B}| \prec\prec |\mathfrak{B}'|$.

If such $\mathfrak{B}'$ exists then, as for categorical algebras, it follows that *any* primal conversion of $\mathfrak{B}$ is linked to $\mathfrak{B}$.

Bibliography

[1] Foster, A. L.: Generalized Boolean theory of universal algebras. I. Math. Z. 58, 306—336 (1953). — [2] Foster, A. L.: Generalized Boolean theory of universal algebras. II. Math. Z. 59, 191—199 (1953). — [3] Foster, A. L.: The identities of — and unique subdirect factorization within — classes of universal algebras. Math. Z. 62, 171—188 (1955).

(Received January 19, 1959)

Hofmann, K. H.
Math. Annalen 138, 239—258 (1959)

Topologische Doppelloops und topologische Halbgruppen

Von

Karl Heinrich Hofmann in Tübingen

Einleitung

In einer früheren Arbeit habe ich die endlichdimensionalen und lokalkompakten topologischen Neokörper klassifiziert [4]; mit diesem Namen bezeichnet man solche topologisch-algebraischen Strukturen, die sich von einem zusammenhängenden lokalkompakten topologischen Schiefkörper nur dadurch unterscheiden, daß von der Addition weder die Assoziativität noch die Kommutativität, sondern nur die eindeutige und stetige Auflösbarkeit der Gleichungen $a + x = y + a = b$ und die Existenz eines neutralen Elementes vorausgesetzt wird. In der vorliegenden Arbeit lassen wir auch noch die beiden Distributivgesetze fallen und beschreiben die multiplikative Struktur zusammenhängender lokalkompakter topologischer Doppelloops mit assoziativer Multiplikation. Zu diesem Zweck entwickeln wir zunächst die Theorie einer speziellen Klasse topologischer Halbgruppen, die eine überall dichte topologische Untergruppe enthalten. Und zwar gehen wir aus von der folgenden Definition:

Eine topologische Halbgruppe K heiße genau dann punktierbar, wenn sie nichtdiskret und homogen ist — d. h. wenn für je zwei Punkte homöomorphe Umgebungen existieren — und wenn sie eine in ihr dichte und offene topologische Untergruppe G enthält, so daß das Komplement $K - G$ nicht leer und total unzusammenhängend ist.

Die Kennzeichnung aller zusammenhängenden und lokalkompakten Halbgruppen dieser Art ergibt sich dann aus den folgenden Hauptsätzen:

Satz I. *K sei eine endlichdimensionale, zusammenhängende, lokalkompakte, punktierbare topologische Halbgruppe. Dann ist K isomorph entweder zur multiplikativen Halbgruppe aller reellen Zahlen, oder zur multiplikativen Halbgruppe aller komplexen Zahlen, oder zur multiplikativen Halbgruppe aller Quaternionen.*

Satz II. *K sei eine unendlichdimensionale, zusammenhängende, lokalkompakte, punktierbare topologische Halbgruppe. Dann gibt es ein Element 0 in K mit $0x = x0 = 0$ für alle $x \in K$, und $K' = K - \{0\}$ ist eine Gruppe mit einer zur additiven Gruppe aller reellen Zahlen isomorphen Einparameteruntergruppe E und einer abzählbaren Menge von normalen, einfachen und einfach zusammenhängenden Lieuntergruppen L_i derart, daß für das in K' normale direkte Produkt B der L_i die Beziehung $K' = EB$ gilt. K' ist homöomorph zu dem Produktraum $E \times B$.*

Offen bleibt hier die Frage, ob solche Halbgruppen, wie sie in Satz II beschrieben sind, überhaupt existieren.

Aus diesen Sätzen folgen dann mühelos die gewünschten Ergebnisse über topologische Doppelloops mit assoziativer Multiplikation:

Satz I'. $(K, +, \cdot)$ *sei eine endlichdimensionale, zusammenhängende, lokalkompakte, topologische Doppelloop mit assoziativer Multiplikation. Dann ist die multiplikative Struktur* $(K, \cdot)$ *von* K *isomorph zur multiplikativen Struktur entweder der reellen Zahlen, oder der komplexen Zahlen, oder der Quaternionen.*

Satz II'. $(K, +, \cdot)$ *sei eine unendlichdimensionale, zusammenhängende, lokalkompakte topologische Doppelloop mit assoziativer Multiplikation. Dann enthält die Gruppe* K', *die durch die Multiplikation auf dem Komplement der Menge aus dem Neutralelement* 0 *der Addition erklärt ist, eine zur additiven Gruppe aller reellen Zahlen isomorphe Einparameteruntergruppe* E *und eine abzählbare Menge von normalen, einfachen und einfach zusammenhängenden Lieuntergruppen* L_i *derart, daß für das in* K' *normale direkte Produkt* B *der* L_i *die Beziehung* $K' = E\,B$ *gilt.* $K' = K - \{0\}$ *ist homöomorph zu dem Produktraum* $E \times B$.

Beim Beweis der Sätze I' und II' aus den Sätzen I und II benötigt man noch nicht einmal die volle Doppelloopeigenschaft; die Sätze bleiben richtig, wenn K eine topologisch algebraische Struktur bezeichnet, auf der überall eine „Addition" $(a, b) \to a + b$ erklärt ist, bezüglich der man stets eine eindeutig und stetig bestimmbare Lösung der sämtlichen Gleichungen $a + x = b$ findet, und in der weiterhin eine überall erklärte „Multiplikation" $(a, b) \to ab$ auf dem Komplement eines einzigen Ausnahmeelements 0 eine Gruppe definiert. Auch hier bleibt die Frage ungeklärt, ob unendlichdimensionale, zusammenhängende, lokalkompakte topologische Doppelloops mit assoziativer Multiplikation wirklich existieren.

Im Fall von vier Dimensionen gehen wir näher ein auf die Struktur solcher topologischer Doppelloops mit assoziativer Multiplikation, in denen wenigstens ein Distributivgesetz gilt; die dabei gewonnenen und in Satz III im Paragraphen 4 zusammengefaßten Ergebnisse sind in gewisser Weise Verallgemeinerungen eines Sachverhalts, der von Kalscheuer bei der Bestimmung aller stetigen Fastkörper angetroffen wurde [5]. Die dort aufgeworfene Frage nach einer abstrakten, also von Darstellungstheorie freien Behandlung der topologischen lokalkompakten Fastkörper über dem Körper der reellen Zahlen ist durch unser Verfahren weitgehend beantwortet.

§ 1. Hilfssätze aus der Gruppentheorie

In den ersten Paragraphen stellen wir eine Serie von Hilfssätzen zusammen, auf die wir dann im Beweis der Hauptsätze zurückgreifen können.

(1.1). *G sei eine zusammenhängende, einfach zusammenhängende lokalkompakte topologische Gruppe. Dann enthält G eine gewisse Zahl n von Einparametergruppen, die wir mit $E_1, \ldots, E_n$ bezeichnen, und die bei den Isomorphismen $f_1, \ldots, f_n$ zur additiven Gruppe R der reellen Zahlen isomorph sind, und ferner eine maximale kompakte Untergruppe B, die ein direktes Produkt einfacher und einfach zusammenhängender Liegruppen ist; die Abbildung $(r_1, \ldots, r_n, x) \to f(r_1) \ldots f(r_n)x$ ist dabei ein Homöomorphismus von $R^n \times B$ auf G. Hat G endliche Dimension, so sind G und B Liegruppen.*

Beweis. Nach dem Satz von IWASAWA (vgl. z. B. [7] S. 188) haben wir nur noch die Struktur einer maximalen kompakten Untergruppe anzugeben. B ist ein Produkt TH einer zentralen Torusgruppe T mit einem projektiven Limes H von halbeinfachen Liegruppen derart, daß TH isomorph ist der Faktorgruppe von $T \times H$ nach einer Gruppe isomorph $T \cap H$. Diese letzte Gruppe ist abgeschlossen und nicht dicht in H oder T. Angenommen nun, T enthielte wenigstens eine nichttriviale eindimensionale abgeschlossene Torusuntergruppe T_0, und T wäre also gleich $T_0 \times T_1$ mit einem komplementären Torusfaktor T_1. Dann wäre $TH = B = T_0 \cdot T_1 B$ und diese Gruppe wäre isomorph einer Faktorgruppe von $T_0 \times T_1 B$ nach einer abgeschlossenen zentralen diskreten Untergruppe isomorph zu $T_0 \cap T_1 B$ ([13] S. 21). Da $T_0 \times T_1 B$ sicher nicht einfach zusammenhängend ist, kann dies bei einer Faktorgruppe nach einem diskreten Normalteiler sicher auch nicht der Fall sein. Also ist notwendig $B = H$. Dann ist aber B ein direktes Produkt einfacher und einfach zusammenhängender Liegruppen ([13] S. 90—91). Die Endlichkeit der Dimension zieht bekanntermaßen die Analytizität der Gruppe nach sich.

(1.2). *G sei eine n-dimensionale Liegruppe. Es gebe eine abgeschlossene Einparameteruntergruppe E und eine abgeschlossene Untergruppe B derart, daß $E \cap B$ diskret ist und $G = EB$ gilt. Dann ist die natürliche Abbildung der Mannigfaltigkeit $E \times B$ in die Mannigfaltigkeit $G = EB$ lokal eineindeutig; insbesondere haben beide Mannigfaltigkeiten die gleiche Dimension.*

Beweis. Wir beweisen die Behauptung dadurch, daß wir zeigen, daß die analytische Abbildung $(x, y) \to xy$ von $E \times B$ in $G = EB$ überall eine eineindeutige lineare Abbildung der Tangentialräume von Urbildpunkt und Bildpunkt erzeugt. Bezeichnet d_a das Differential der Abbildung $z \to az$ und δ_a das Differential der Abbildung $z \to za$, ist ferner $\mathfrak{x}$ ein Tangentialvektor von E in 1 und $\mathfrak{y}$ ein Tangentialvektor von B in 1, dann ist $(d_x \mathfrak{x}, d_y \mathfrak{y})$ ein Tangentialvektor der Mannigfaltigkeit $E \times B$ im Punkt (x, y). Sein Bild im Tangentialraum von G im Punkt xy bei der durch die Abbildung $(x, y) \to xy$ bewirkten Abbildung der Tangentialräume ist der Tangentialvektor $d_x \mathfrak{x} \delta_y + d_x d_y \mathfrak{y}$ $= d_x d_y (ad(y^{-1}) \mathfrak{x} + \mathfrak{y})$, wenn $z \to ad(z)$ die adjungierte Darstellung der Gruppe G auf ihrer Infinitesimalalgebra bezeichnet, d. h. die Abbildung, die jedem Element z aus G die durch den inneren Automorphismus $u \to zuz^{-1}$ von G bewirkte Transformation der Infinitesimalalgebra von G zuordnet. Dieser Tangentialvektor ist aber dann und nur dann gleich Null, wenn $\mathfrak{x} + ad(y)\mathfrak{y} = 0$; da aber $ad(y)\mathfrak{y}$ ein Tangentialvektor von B in 1 ist und die Vektorräume der zu E und B gehörigen Lieunteralgebren eine direkte Summe bilden, ist dies genau dann der Fall, wenn $\mathfrak{x} = 0$ und $\mathfrak{y} = 0$. Dann ist aber die lineare Abbildung der Tangentialvektorräume sicher eineindeutig. Vgl. auch [11] S. 22—90.

(1.3). *G sei eine n-dimensionale Liegruppe; es gebe eine abgeschlossene nicht kompakte Einparameteruntergruppe E und eine kompakte zusammenhängende Untergruppe B in G derart, daß $G = EB$. Dann ist B normal in G.*

Beweis. B ist im Falle $n > 1$, der hier allein interessiert, eine $n-1$-dimensionale Liegruppe; denn da E nicht kompakt ist, gilt $E \cap B = 1$ und nach (1.2) hat $E \times B$ die gleiche Dimension wie EB. Durch ein Element $x \in G - B$

laufe die Einparametergruppe F. Ist $\overline{F}$ kompakt, also eine Abelsche Liegruppe und daher ein endlichdimensionaler Torus, dann gibt es in $\overline{F}$ eine abgeschlossene Einparametergruppe F', die nicht in B enthalten ist und als kompakte Gruppe auch in G abgeschlossen ist. $F' \cap B$ ist diskret; also ist nach (1.2) $F'B$ eine n-dimensionale Teilmannigfaltigkeit von G und stimmt daher mit G überein. Nun ist aber $F'B$ als Komplexprodukt zweier kompakter Teilräume kompakt; aber G ist nicht kompakt. Daraus folgt, daß $\overline{F}$ nicht kompakt sein kann. x liegt also nicht in einer kompakten Untergruppe von G. Das hat zur Folge, daß ein Element $x \in G$ genau dann in einer kompakten Untergruppe liegt, wenn $x \in B$. Da die Menge aller Elemente, die in einer kompakten Untergruppe von G liegen, bei allen Automorphismen von G in sich abgebildet wird, ist B eine charakteristische Untergruppe und daher normal.

(1.4). *G sei eine lokal kompakte zusammenhängende topologische Gruppe, deren homogener Raum nach einer kompakten maximalen Untergruppe B die Dimension* 1 *habe. In der Bezeichnungsweise von* (1.1) *ist also $n = 1$. Dann ist B eindeutig bestimmt und ist normal in G. Ist G einfach zusammenhängend und sind $L_i (i \in I)$ die nach* (1.1) *existierenden einfachen und einfach zusammenhängenden direkten Faktoren von B, so sind auch diese normal in G.*

Beweis. In jeder kompakten Umgebung U von 1 gibt es einen kompakten, in B enthaltenen Normalteiler H_U derart, daß G/H_U eine Liegruppe ist: Es gibt jedenfalls einen Normalteiler H_U in U, so daß G/H_U eine Liegruppe ist ([7] Chap. IV). Die Untergruppe BH_U ist kompakt in G, umfaßt B und muß der Maximalität von B wegen mit B übereinstimmen. Also ist H_U in B enthalten. Nun ist $G/H_U = EB/H_U$ mit $E_1 = E$ eine Liegruppe von dem in (1.3) beschriebenen Typ. B/H_U ist also normal in G/H_U, d. h. für alle $x \in G$ gilt $xH_U \cdot BH_U = BH_U \cdot xH_U$. Daraus folgt $xB = \bigcap_U xH_U \cdot BH_U = \bigcap_U BH_U \cdot xH_U = Bx$, d. h. B ist normal in G. Nun sei α ein beliebiger Automorphismus von B. Dieser kann in dem direkten Produkt einfacher nicht Abelscher Liegruppen L_i allenfalls die Faktoren vertauschen und auf den einzelnen Faktoren einen Automorphismus induzieren. Nun sei E wieder eine Einparametergruppe von G, für die $G = EB$ gilt. Es sei $l \in L_{i_0}$ ein von 1 verschiedener Punkt aus einem der direkten Faktoren. Das Bild von E bei der Abbildung $x \to x^{-1}lx$ ist zusammenhängend und enthält wegen $l \neq 1$ nicht den Punkt 1. Es ist andrerseits in $\bigcup_{i \in I} L_i$ enthalten. Da zwei verschiedene der direkten Faktoren nur den Punkt 1 gemeinsam haben, muß dieses Bild ganz in L_{i_0} enthalten sein. Alle L_i sind also mit E vertauschbar, was zur Normalität von L_i in $G = EB$ ausreicht.

§ 2. Vierdimensionale einfach zusammenhängende Liegruppen mit dreidimensionaler kompakter Untergruppe

(2.1). *G sei eine vierdimensionale Liegruppe, die einen einfach zusammenhängenden Normalteiler B enthält, der zur einfach zusammenhängenden Überlagerungsgruppe der Gruppe $SO(3)$ aller orthogonalen Transformationen des dreidimensionalen euklidischen Raums mit der Determinante* 1, *d. h. zur multi-*

plikativen Gruppe der Quaternionen mit der Norm 1 isomorph ist. Dann hat die Infinitesimalalgebra $\mathfrak{G}$ von G eine Basis $\mathfrak{l}_0, \mathfrak{l}_1, \mathfrak{l}_2, \mathfrak{l}_3$ mit einer Multiplikationstafel der Form

	$\mathfrak{l}_0$	$\mathfrak{l}_1$	$\mathfrak{l}_2$	$\mathfrak{l}_3$
$\mathfrak{l}_0$	0	0	0	0
$\mathfrak{l}_1$	0	0	$\mathfrak{l}_3$	$-\mathfrak{l}_2$
$\mathfrak{l}_2$	0	$-\mathfrak{l}_3$	0	$\mathfrak{l}_1$
$\mathfrak{l}_3$	0	$\mathfrak{l}_2$	$-\mathfrak{l}_1$	0

Ist G einfach zusammenhängend, so ist G isomorph zur multiplikativen Gruppe des Schiefkörpers der Quaternionen.

Beweis. Die Lie-Algebra $\mathfrak{G}$ von G ist eine direkte Summe aus einer eindimensionalen, von der infinitesimalen Transformation $\mathfrak{l}_0'$ erzeugten Unteralgebra $\mathfrak{A}$ und dem zu B gehörenden dreidimensionalen Ideal $\mathfrak{B}$. Die Algebra $\mathfrak{B}$ ist isomorph zu der Algebra, die entsteht, wenn man den dreidimensionalen Vektorraum R^3 über dem Körper R der reellen Zahlen mit dem Vektorprodukt $(\mathfrak{x}, \mathfrak{y}) \to \mathfrak{x} \times \mathfrak{y}$ ausstattet; um das einzusehen, braucht man nur die üblicherweise verwendete Basis der Infinitesimalalgebra der Drehgruppe und ihre Multiplikationstafel zu betrachten. Die Abbildung $\mathfrak{x} \to [\mathfrak{l}_0', \mathfrak{x}]$ mit der Liemultiplikation $(\mathfrak{x}, \mathfrak{y}) \to [\mathfrak{x}, \mathfrak{y}]$ bildet in $\mathfrak{G}$ das Ideal $\mathfrak{B}$ auf sich ab. Auf $\mathfrak{B}$ stellt sie eine Ableitung dar, d. h. eine lineare Abbildung D mit der Eigenschaft $D[\mathfrak{x}, \mathfrak{y}] = [D\mathfrak{x}, \mathfrak{y}] + [\mathfrak{x}, D\mathfrak{y}]$; dies ist eine Folge der Jacobi-Identität. Es ist nun eine Aufgabe der elementaren analytischen Geometrie, festzustellen, daß es in R^3 zu jeder linearen Abbildung A eine und nur eine lineare Abbildung $\bar{A}$ gibt mit $\bar{A}(\mathfrak{x} \times \mathfrak{y}) = A\mathfrak{x} \times \mathfrak{y} + \mathfrak{x} \times A\mathfrak{y}$; und $\bar{A}$ hat dann die Form $(\text{Spur } A) \cdot 1 - A^T$ mit der identischen Abbildung 1 und der transponierten Abbildung A^T. Aus $A = (\text{Spur } A) \cdot 1 - A^T$ folgt dann aber Spur $A = 0$ und damit $A = A^T$, d. h. A ist schiefsymmetrisch, und es gibt genau einen Vektor $\mathfrak{a}$ so, daß $A\mathfrak{x} = \mathfrak{a} \times \mathfrak{x}$. Ist $\mathfrak{a} \neq 0$, dann wählen wir den Einheitsvektor $\mathfrak{e}_1$ in der von $\mathfrak{a}$ aufgespannten Vektorschar und ergänzen ihn zu einem orthogonalen Dreibein $\mathfrak{e}_1, \mathfrak{e}_2, \mathfrak{e}_3$, für das $\mathfrak{e}_n \times \mathfrak{e}_{n+1} = \mathfrak{e}_{n+2}$ gilt, wenn die Indizes modulo 3 gerechnet werden. Falls $\mathfrak{a} = 0$ ist, wählen wir den Vektor $\mathfrak{e}_1$ beliebig und ergänzen zu einem Dreibein wie eben angegeben. Es gibt also eine reelle Zahl λ so, daß $A\mathfrak{x} = \lambda(\mathfrak{e}_1 \times \mathfrak{x})$; und in N existiert eine Basis $\mathfrak{l}_1, \mathfrak{l}_2, \mathfrak{l}_3$ mit der Eigenschaft, daß $[\mathfrak{l}_0', \mathfrak{x}] = \lambda[\mathfrak{l}_1, \mathfrak{x}]$ auf $\mathfrak{B}$ gilt und die Basis $\mathfrak{l}_0', \mathfrak{l}_1, \mathfrak{l}_2, \mathfrak{l}_3$ eine Multiplikationstafel der folgenden Form hat:

	$\mathfrak{l}_0'$	$\mathfrak{l}_1$	$\mathfrak{l}_2$	$\mathfrak{l}_3$
$\mathfrak{l}_0'$	0	0	$\lambda\mathfrak{l}_3$	$-\lambda\mathfrak{l}_2$
$\mathfrak{l}_1$	0	0	$\mathfrak{l}_3$	$-\mathfrak{l}_2$
$\mathfrak{l}_2$	$-\lambda\mathfrak{l}_3$	$-\mathfrak{l}_3$	0	$\mathfrak{l}_1$
$\mathfrak{l}_3$	$\lambda\mathfrak{l}_2$	$\mathfrak{l}_2$	$-\mathfrak{l}_1$	0

Ist nun $\lambda \neq 0$, so setzen wir $\mathfrak{l}_0 = \mathfrak{l}_1 - \lambda^{-1}\mathfrak{l}_0$. Dann spannen $\mathfrak{l}_0, \mathfrak{l}_1, \mathfrak{l}_2, \mathfrak{l}_3$ wieder ganz $\mathfrak{G}$ auf und es gilt jetzt $[\mathfrak{l}_0, \mathfrak{l}_n] = 0$ für alle n. Das Zentrum von G wird durch $\mathfrak{l}_0$ erzeugt. Der Nachweis, daß G im Falle des einfachen Zusammenhangs zur multiplikativen Gruppe des Schiefkörpers der Quaternionen isomorph ist, wird später mühelos daraus folgen, daß alle vierdimensionalen Doppelloops — eine Klasse später zu definierender topologisch algebraischer Strukturen, der der Quaternionenschiefkörper angehört — unter der Voraussetzung der Assoziativität der Multiplikation die nämliche Multiplikationsgruppe haben, die den obengenannten Voraussetzungen genügt.

(2.2). *G sei eine einfach zusammenhängende vierdimensionale nicht kompakte Liegruppe mit einer dreidimensionalen kompakten Untergruppe B. Dann ist G isomorph zur multiplikativen Gruppe der von Null verschiedenen Quaternionen.*

Beweis. Nach (1.1) und (1.3) ist B einfach zusammenhängend und normal. Die einzige dreidimensionale kompakte und einfach zusammenhängende Liegruppe ist die einfach zusammenhängende Überlagerungsgruppe der $SO(3)$; denn jede kompakte Liegruppe ist Faktorgruppe eines direkten Produktes aus einer zentralen Torusgruppe und einer halbeinfachen Untergruppe nach einer diskreten Gruppe; sie ist also nur dann einfach zusammenhängend, wenn die zentrale zusammenhängende Untergruppe diskret ist, d. h. die Gruppe halbeinfach ist. Die einfachen nichtabelschen Liegruppen sind aber bekannt; sie sind mindestens dreidimensional und bei der genauen Dimension drei isomorph zu der $SO(3)$ oder ihrer einfach zusammenhängenden, zu der 3-Sphäre homöomorphen Überlagerungsgruppe.

(2.3). *G sei eine vierdimensionale, zur multiplikativen Gruppe des Schiefkörpers der Quaternionen isomorphe Liegruppe mit der Infinitesimalalgebra $\mathfrak{G}$. Sie ist also das eindeutig bestimmte direkte Produkt einer zur additiven Gruppe R der reellen Zahlen isomorphen Einparametergruppe E und einer dreidimensionalen Sphärengruppe B mit dem zugehörigen Ideal $\mathfrak{B}$ von $\mathfrak{G}$. Die Basis $\mathfrak{l}_0, \mathfrak{l}_1, \mathfrak{l}_2, \mathfrak{l}_3$ sei nach* (2.1) *aufgestellt. Jeder Homomorphismus h von G in die von $\mathfrak{l}_1$ erzeugte zu einem eindimensionalen Torus isomorphe Untergruppe T ist dadurch gekennzeichnet, daß in einer hinreichend kleinen Umgebung von* 1 *mit kanonischen Koordinaten erster Art $x_0, \ldots, x_3$ einem Element $x = \exp \sum_{n=0}^{3} x_n \mathfrak{l}_n$ das Element $h(x) = \exp(-\lambda x_0 \mathfrak{l}_1)$ mit einer reellen Zahl λ zugeordnet wird. Durch die Festsetzung $(x, y) \to x \circ y = h(y)^{-1} x h(y) y$ wird auf G eine neue Verknüpfung erklärt, bezüglich der $(G, \circ)$ eine (im übrigen wegen* (2.2) *zu $(G, .)$ isomorphe) Liegruppe wird, und der in der Liealgebra $\mathfrak{G}$ eine neue Liemultiplikation $(\mathfrak{x}, \mathfrak{y}) \to [\mathfrak{x}, \mathfrak{y}]_\circ$ entspricht, die durch $[\mathfrak{x}, \mathfrak{y}]_\circ = [L\mathfrak{x}, L\mathfrak{y}]$ mit einer linearen Abbildung L des Vektorraums der Liealgebra $\mathfrak{G}$ erklärt ist; L hat bezüglich der Basis $\mathfrak{l}_0, \mathfrak{l}_1, \mathfrak{l}_2, \mathfrak{l}_3$ die Matrix*

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ \lambda & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Diese Basis hat bezüglich der neuen Verknüpfung die Multiplikationstafel

$\circ$	$\mathfrak{l}_0$,	$\mathfrak{l}_1$,	$\mathfrak{l}_2$,	$\mathfrak{l}_3$
$\mathfrak{l}_0$	0	0	$\lambda\mathfrak{l}_3$	$-\lambda\mathfrak{l}_2$
$\mathfrak{l}_1$	0	0	$\mathfrak{l}_3$	$-\mathfrak{l}_2$
$\mathfrak{l}_2$	$-\lambda\mathfrak{l}_3$	$-\mathfrak{l}_3$	0	$\mathfrak{l}_1$
$\mathfrak{l}_3$	$\lambda\mathfrak{l}_2$	$\mathfrak{l}_2$	$-\mathfrak{l}_1$	0

Beweis. Jeder Homomorphismus h von G in T hat einen Kern, der B umfaßt; denn B ist eine einfache Liegruppe und T ist abelsch. Falls h nicht ganz G auf 1 abbildet, ist der Kern ein direktes Produkt einer diskreten Untergruppe D von E mit B; denn da E nicht ganz im Kern liegen kann, ist der Schnitt des Kerns mit E eine diskrete Untergruppe D von E und DB ist notwendig der volle Kern, da mit xb (wo $x \in E$, $b \in B$) zugleich auch $xbb^{-1} = x$ im Kern liegt. Umgekehrt gibt es zu jeder diskreten Untergruppe D von E einen Homomorphismus h von G auf T: Bezeichnet nämlich p die Projektion von EB auf E und gibt t einen Homomorphismus von E auf T mit Kern D an, so ist $t \circ p$ der gesuchte Homomorphismus. In einer genügend kleinen Umgebung von 1 hat dieser Homomorphismus mit geeigneten Koordinaten die Form $h(\exp \sum_{n=0}^{3} x_n \mathfrak{l}_n) = \exp(-\lambda x_0 \mathfrak{l}_1)$ und λ kann jeden reellen Wert annehmen. Der Fall $\lambda = 0$ liefert den trivialen Homomorphismus. Das Minuszeichen führen wir nur zur Vereinfachung der nachfolgenden Rechnungen ein. Wir setzen der Bequemlichkeit halber $h(y)^{-1} x h(y) = x^y$. Dann gilt $(x^y)^z = x^{yz} = x^{zy}$, $(xy)^z = x^z y^z$ und $x^{y^z} = x^y$, letzteres wegen $h(y^z) = h(h(z)^{-1} y h(z)) = h(y)$. Die neue Verknüpfung $x \circ y = x^y y$ ist assoziativ: $(x \circ y) \circ z = (x^y y)^z z = x^{yz} y^z z$, andrerseits ist $x \circ (y \circ z) = x \circ (y^z z) = x^{y^z z} y^z z = x^{yz} y^z z$. Sie besitzt ein neutrales Element, nämlich 1 und gestattet zu jedem Element x die Bildung des Inversen $\bar{x}$, nämlich $\bar{x} = (x^{-1})^{x^{-1}}$; denn

$$\bar{x} \circ x = (x^{-1})^{x^{-1}x} x = x^{-1} x = 1$$

und

$$x \circ \bar{x} = x^{(x^{-1})^{x^{-1}}} (x^{-1})^{x^{-1}} = x^{x^{-1}} (x^{-1})^{x^{-1}} = 1 .$$

$(G, \circ)$ ist also eine Gruppe und sogar eine vierdimensionale topologische und analytische Gruppe; sie ist daher isomorph zu $(G, .)$; dabei ist $xy = x \cdot y$ gesetzt, so daß der Punkt für die alte Multiplikation steht. Wir untersuchen nun, welche neue Liemultiplikation in $\mathfrak{G}$ die Folge des Übergangs zu einer neuen Multiplikation in G ist. Die Basis $\mathfrak{l}_0, \mathfrak{l}_1, \mathfrak{l}_2, \mathfrak{l}_3$ von $\mathfrak{G}$ sei wie in (2.1) gewählt. Der Kommutator $(u, v) = uvu^{-1}v^{-1}$ in $(G, .)$ erlaubt es, die neue Multiplikation in G wie folgt darzustellen: $x \circ y = (h(y)^{-1}, x)\, xy$. Mit Hilfe der Formel von Campbell-Hausdorff berechnen wir den Kommutator $x \circ y \circ \overline{y \circ x}$ bezüglich der neuen Verknüpfung. Zunächst gilt (immer für ausreichend kleine Werte von t) die Beziehung

$$x \circ y = (\exp \lambda t\, y_0 \mathfrak{l}_1, \exp t \mathfrak{x}) \exp t \mathfrak{x} \exp t \mathfrak{y}$$

$$= \exp\Big(t(\mathfrak{x} + \mathfrak{y}) + \frac{t^2}{2}[\mathfrak{x}, \mathfrak{y}] + t^2[\mathfrak{x}, \lambda y_0 \mathfrak{l}_1] + 0(t^3)\Big) = \exp \mathfrak{u} \, ,$$

wobei an Stelle des Ausdrucks $0(t^3)$ höchstens Glieder stehen, in denen t mindestens in der dritten Potenz vorkommt. Analog gilt

$$y \circ x = \exp\left(t(\mathfrak{x} + \mathfrak{y}) + \frac{t^2}{2}[\mathfrak{y}, \mathfrak{x}] + t^2[\mathfrak{y}, \lambda x_0 \mathfrak{l}_1] + 0(t^3)\right) = \exp \mathfrak{v}\,.$$

Damit berechnen wir nun

$$\begin{aligned}
\overline{y \circ x} &= \overline{\exp \mathfrak{v}} = (\exp - \mathfrak{v})^{\exp - \mathfrak{v}} = h(\exp \mathfrak{v})\,(\exp - \mathfrak{v})\,h(\exp \mathfrak{v})^{-1} \\
&= \exp(-t\lambda(x_0 + y_0)\,\mathfrak{l}_1 + 0(t^2))\,(\exp - \mathfrak{v})\,\exp(t\lambda(x_0 + y_0)\,\mathfrak{l}_1 + 0(t^2)) \\
&= \exp(-\mathfrak{v} + [t\lambda(x_0 + y_0)\,\mathfrak{l}_1, -\mathfrak{v}] + 0(t^3)) \\
&= \exp\left(-t(\mathfrak{x}+\mathfrak{y}) + t^2\left(\frac{1}{2}[\mathfrak{x},\mathfrak{y}] - [\mathfrak{y}, \lambda x_0 \mathfrak{l}_1] - [\lambda(x_0 + y_0)\mathfrak{l}_1, \mathfrak{x}+\mathfrak{y}]\right) + 0(t^3)\right).
\end{aligned}$$

Weiterhin bekommt man

$$\begin{aligned}
\exp \mathfrak{u}^{\overline{\exp \mathfrak{v}}} &= \exp \mathfrak{u}^{\exp - \mathfrak{v}} = h(\exp \mathfrak{v})\,\exp \mathfrak{u}\, h(\exp \mathfrak{v})^{-1} \\
&= \exp(-t\lambda(x_0 + y_0)\,\mathfrak{l}_1 + 0(t^2))\,\exp \mathfrak{u}\,\exp(t\lambda(x_0 + y_0)\mathfrak{l}_1 + 0(t^2)) \\
&= \exp(\mathfrak{u} + [t\lambda(x_0 + y_0)\mathfrak{l}_1, \mathfrak{u}] + 0(t^3)) \\
&= \exp\left(t(\mathfrak{x} + \mathfrak{y}) + t^2\left(\frac{1}{2}[\mathfrak{x}, \mathfrak{y}] + [\mathfrak{x}, \lambda y_0 \mathfrak{l}_1] + [\lambda(x_0 + y_0)\mathfrak{l}_1, \mathfrak{x}+\mathfrak{y}]\right) + 0(t^3)\right).
\end{aligned}$$

Daraus gewinnen wir schließlich

$$\begin{aligned}
x \circ y \circ \overline{y \circ x} &= \exp \mathfrak{u}^{\exp - \mathfrak{v}}\, \overline{\exp \mathfrak{v}} \\
&= \exp(t^2([\mathfrak{x}, \mathfrak{y}] + [\mathfrak{x}, \lambda y_0 \mathfrak{l}_1] + [\lambda x_0 \mathfrak{l}_1, \mathfrak{y}]) + 0(t^3)) \\
&= \exp(t^2[\mathfrak{x} + \lambda x_0 \mathfrak{l}_1, \mathfrak{y} + \lambda y_0 \mathfrak{l}_1] + 0(t^3)) \quad \text{wegen } [\mathfrak{l}_1, \mathfrak{l}_1] = 0.
\end{aligned}$$

Wir bezeichnen nun mit L diejenige lineare Abbildung der Liealgebra $\mathfrak{G}$ in sich, die bezüglich der Basis $\mathfrak{l}_0, \mathfrak{l}_1, \mathfrak{l}_2, \mathfrak{l}_3$ die Matrix

$$\begin{bmatrix} 1000 \\ \lambda 100 \\ 0010 \\ 0001 \end{bmatrix}$$

besitzt.

Durch die Festsetzung $[\mathfrak{x}, \mathfrak{y}]_\circ = [L\mathfrak{x}, L\mathfrak{y}]$ wird auf $\mathfrak{G}$ eine neue Liemultiplikation erklärt, die wieder schiefsymmetrisch ist und der Jacobiidentität genügt. Außerdem ist nach dem oben Bewiesenen $(\mathfrak{G}, +, [,]_\circ)$ die zu der Liegruppe $(K', \circ)$ gehörende Liealgebra. Die Multiplikationstafel von $\mathfrak{G}$ für die Basis $\mathfrak{l}_0, \mathfrak{l}_1, \mathfrak{l}_2, \mathfrak{l}_3$ lautet bezüglich der neuen Multiplikation

	$\mathfrak{l}_0$	$\mathfrak{l}_1$	$\mathfrak{l}_2$	$\mathfrak{l}_3$
$\mathfrak{l}_0$	0	0	$\lambda\mathfrak{l}_3$	$-\lambda\mathfrak{l}_2$
$\mathfrak{l}_1$	0	0	$\mathfrak{l}_3$	$-\mathfrak{l}_2$
$\mathfrak{l}_2$	$-\lambda\mathfrak{l}_3$	$-\mathfrak{l}_3$	0	$\mathfrak{l}_1$
$\mathfrak{l}_3$	$-\lambda\mathfrak{l}_2$	$\mathfrak{l}_2$.	$-\mathfrak{l}_1$	0

§ 3. Punktierbare topologische Halbgruppen

(3.1). *Eine topologische Halbgruppe K ist eine Halbgruppe mit einer assoziativen Verknüpfung $(x, y) \to xy$, die bezüglich einer außerdem auf K erklärten Topologie stetig ist. Ein Ideal J in K ist eine Unterstruktur von K mit der Eigenschaft, daß aus $j \in J$ stets $xj \in J$ und $jx \in J$ folgt für alle $x \in K$.*

(3.2). *Ist J ein total unzusammenhängendes Ideal in einer zusammenhängenden topologischen Halbgruppe K, die ein neutrales Element 1 mit $1x = x1 = x$ für alle $x \in K$ enthält, so besteht J nur aus einem Punkt.*

Beweis. Es sei $j \in J$. Die Abbildung $x \to xj$ bildet den zusammenhängenden Raum K in den total unzusammenhängenden Raum J ab und ist daher konstant. Da das Bild das Element $1j = j$ enthält, gilt $xj = j$ für alle $x \in K$. Da das Ideal symmetrisch definiert ist, gilt ebenso $jx = j$ für alle $x \in K$. Sind nun insbesondere j_1 und j_2 zwei Punkte aus J, so gilt $j_1 = j_1 j_2 = j_2$, d. h. J enthält nur einen Punkt.

(3.3). *Wir nennen eine topologische Halbgruppe K dann und nur dann punktierbar, wenn sie folgende Eigenschaften aufweist:*

a) K ist homogen und nichtdiskret. Wir nennen dabei einen Raum homogen, wenn zu zwei Punkten immer zwei homöomorphe Umgebungen existieren.

b) Es gibt eine total unzusammenhängende nichtleere nirgends dichte Teilmenge J in K derart, daß $K - J$ eine in K offene Gruppe ist.

Eine topologische punktierbare Halbgruppe mit einer Teilmenge J, die den Bedingungen a) und b) dieses Absatzes genügen, nennen wir auch in J punktierbar; die Menge J bezeichnen wir als eine Ausnahmemenge.

(3.4). *In einer in J punktierbaren topologischen Halbgruppe ist J ein Ideal.*

Beweis. j liege in der Ausnahmemenge J. Die Abbildungen $x \to xj$ und $x \to jx$ bilden dann ganz $K - J$ in J ab. Wäre nämlich z. B. $xj = y \in K - J$ für ein $x \in K - J$, so läge $j = x^{-1}xj = x^{-1}y$ in $K - J$, was nicht der Fall ist. Jetzt sei $x \in J$. Da J in K nirgends dicht ist, gibt es in jeder Umgebung U von j ein z aus $K - J$. Das Element $y^{-1}x$ liegt nach dem oben bewiesenen für $x \in J$ in J und dann ebenso das Element $(y^{-1}x)z$. Da J abgeschlossen ist, liegt der Grenzwert dieser Elemente für $z \to j$ in J; andrerseits ist dieser Grenzwert gleich 1. Aber 1 liegt nicht in J. Daher gilt $xj \in J$ und ebenso $jx \in J$ für *alle* $x \in K$.

(3.5). *Eine zusammenhängende topologische punktierbare Halbgruppe hat eine eindeutig bestimmte Ausnahmemenge, die nur aus einem Punkt besteht.*

Beweis. Die Behauptung folgt direkt aus (3.4), (3.3), (3.2) und der Bemerkung, daß das Neutralelement der dichten Untergruppe stets eine Eins der Halbgruppe ist.

Im folgenden bezeichnen wir stets das Ausnahmeelement einer zusammenhängenden topologischen punktierbaren Halbgruppe mit **0**.

(3.6). *Eine zusammenhängende topologische punktierbare Halbgruppe K ist stets einfach zusammenhängend.*

Beweis. $\tilde{K}$ sei eine einfach zusammenhängende Überlagerungsstruktur von K mit der Überlagerungsabbildung h. Der Teilraum $h^{-1}(0)$ von $\tilde{K}$ ist ein diskretes Ideal in $\tilde{K}$, das nach (3.2) nur aus einem Punkt bestehen darf, denn für $\tilde{K}$ existiert in $h^{-1}(1)$ ein neutrales Element $\tilde{1}$.

Dieser Hilfssatz bedarf einer Anmerkung: Wir haben vorausgesetzt, daß eine einfachzusammenhängende Überlagerungsstruktur existiert. Bei Verwendung des bei CHEVALLEY [1] S. 40ff. angegebenen Überlagerungsbegriffs ist dazu im allgemeinen noch die Voraussetzung des lokalen Zusammenhangs

und des lokal einfachen Zusammenhangs notwendig. Im Fall des von R. S. Novosad [10] S. 216ff. eingeführten Begriffs der Überlagerung und des einfachen Zusammenhangs genügt für kurvenzusammenhängende Räume gerade die angegebene Formulierung. Im übrigen sei auch auf [3] S. 216 und 217 verwiesen, wo der in (3.6) angegebene Sachverhalt mit etwas abweichenden Bezeichnungen ausführlicher dargelegt ist.

(3.7). *K sei eine in einem Punkt 0 punktierbare topologische Halbgruppe. Ist U eine Umgebung von 0 und M ein beliebiger kompakter Teilraum von K, dann gibt es eine Umgebung V von 0 derart, daß VM in U liegt. Ist K lokalkompakt, so besitzt jeder Punkt von K eine abzählbare Basis für seinen Umgebungsfilter.*

Beweis. Angenommen, zu jeder Umgebung V von 0 gäbe es ein Element $x_V \neq 0$ in V, für das ein Element m_V in M so existiert, daß $x_V m_V$ nicht in U liegt. Wenn V den Umgebungsfilter $\mathfrak{U}$ von 0 durchläuft, stellt $(x_V m_V)_{V \in \mathfrak{U}}$ eine (Moore-Smith-) Folge dar, die in $K - U$ verläuft. $(x_V)_{V \in \mathfrak{U}}$ konvergiert gegen 0; da M kompakt ist, können wir o. B. d. A. annehmen, daß $(m_V)_{V \in \mathfrak{U}}$ in M gegen $m \in M$ konvergiert. Dann konvergiert aber $(x_V m_V)_{V \in \mathfrak{U}}$ gegen $0m = 0$; da 0 nicht in $\overline{K - U}$ liegt, haben wir einen Widerspruch gefunden. Ist K lokalkompakt, so können wir für M eine kompakte Umgebung von 0 nehmen und eine abzählbare Folge $(x_n)_{n=1,\ldots}$ finden, die 0 als Adhärenzpunkt hat; denn in einer kompakten Umgebung hat jede Folge einen Adhärenzpunkt und infolge der Homogenität ist auch 0 ein Adhärenzpunkt einer abzählbaren Folge. Zu jeder Umgebung U gibt es eine Umgebung V mit $VM \subsetneq U$; erst recht gilt dann $b_m M \in U$ für ein $x_m \in V$, welches sicher existiert, da 0 Adhärenzpunkt von $(x_n)_{n=1,\ldots}$ ist. Die Mengen $x_n M$ bilden also eine abzählbare Basis für die Umgebungen von 0.

(3.8). K sei eine lokalkompakte, in 0 punktierbare topologische Halbgruppe. Dann ist K in einem homogenen Raum dicht, der außer K nur noch einen weiteren Punkt ∞ enthält. Die Abbildung $x \to x^{-1}$ hat eine Fortsetzung zu einem Homöomorphismus f von $K \cup \{\infty\}$ auf sich mit $f(0) = \infty$ und $f(\infty) = 0$.

(3.9). K sei eine lokalkompakte zusammenhängende topologische in 0 punktierbare Halbgruppe. Dann ist K lokalzusammenhängend.

Die beiden Sätze (3.8) und (3.9) werden hier nicht bewiesen, da sie in einer früheren Arbeit bewiesen wurden: [3], (1.5), (1.6). Sie sind dort formuliert für topologische Doppelloops, deren multiplikative Struktur gerade die in (3.8) und (3.9) vorausgesetzten Eigenschaften hat und sogar dort nicht einmal assoziativ vorausgesetzt ist, während von der Addition nichts in den Beweis einging.

(3.10). *Ist ein euklidischer Raum R^n in n Dimensionen in einem kompakten Raum E dicht, der genau zwei verschiedene Punkte a_1 und a_2 außerhalb R^n enthält, so ist $n = 1$, und E ist ein abgeschlossenes Intervall mit den Endpunkten a_1 und a_2. Dabei nennen wir einen euklidischen Raum jeden topologischen Raum, der zu einem euklidischen homöomorph ist, und benützen den Kompaktheitsbegriff im Sinne von Bourbaki, bei dem immer das Hausdorffsche Trennungsaxiom inbegriffen ist.*

Beweis. ϱ sei diejenige Äquivalenzrelation auf E, die a_1 und a_2 identifiziert. Da die Einpunktkompaktifizierung $R^n \cup \{\infty\}$ von R^n bis auf Homöomorphie eindeutig bestimmt und einer n-Sphäre S^n topologisch äquivalent ist, ist damit auch E/ϱ zu S^n homöomorph. Es seien U_1 und U_2 punktfremde Umgebungen von a_1 bzw. a_2. Wenn mit $\varrho(x)$ die ϱ-Klasse von x bezeichnet wird, ist $\varrho(U_1) \cup \varrho(U_2)$ eine Umgebung von $\varrho(a_1) = \varrho(a_2) = \infty$, die eine offene n-Zelle V enthält. $\varrho^{-1}(V)$ ist die Vereinigung punktfremder offener Umgebungen V_1 und V_2 von a_1 bzw. a_2. Der Teilraum $V_1 \cup V_2 - (\{a_1\} \cup \{a_2\})$ wird bei $x \to \varrho(x)$ homöomorph auf $V - \{\infty\}$ abgebildet. $V - \{\infty\}$ ist also nicht zusammenhängend, weshalb V notwendig eine 1-Zelle ist. Daher ist $n = 1$.

(3.11). *F sei ein topologischer Raum mit einer Äquivalenzrelation ϱ derart, daß F/ϱ einem euklidischen Raum R^n homöomorph sei. (Dies ist z. B. der Fall, wenn F dem topologischen Produkt von R^n und einem beliebigen topologischen Raum homöomorph ist.) F sei in einem kompakten Raum E dicht, der genau zwei verschiedene Punkte a_1 und a_2 außerhalb F enthält. Dann ist $n = 1$.*

Beweis. Auf E definieren wir eine Äquivalenzrelation $\bar{\varrho}$, deren Klasse aus den Mengen $\{a_1\}$ und $\{a_2\}$ sowie den ϱ-Klassen von F besteht. Der Quotientenraum $E/\bar{\varrho}$ enthält den in ihm dichten Teilraum $\bar{\varrho}(F)$, der zu F/ϱ und daher zu R^n homöomorph ist. Aus (3.10) folgt dann die Behauptung.

(3.12). *Ist K eine lokalkompakte zusammenhängende in 0 punktierbare topologische Halbgruppe, so gibt es in $K - \{0\}$ eine Einparameteruntergruppe E, die zur additiven Gruppe R der reellen Zahlen isomorph ist, sowie einen größten kompakten Normalteiler B derart, daß die Gruppe $K - \{0\}$ das Produkt EB ist und die Abbildung $(x, b) \to xb$ ein Homöomorphismus des Produktraums $E \times B$ auf $EB = K - \{0\}$ ist.*

Beweis. Nach dem in (1.1) zitierten Satz von Iwasawa und dem sich nicht auf den einfachen Zusammenhang beziehenden Teil von (1.4) wird es genügen, festzustellen, daß immer dann, wenn nach dem Satz von Iwasawa n zu R isomorphe Einparametergruppen $E_1, \ldots, E_n$ und eine maximale kompakte Untergruppe B so vorliegen, daß $K - \{0\}$ zu dem Produktraum $E_1 \times \cdots \times E_n \times B$ homöomorph ist, die Zahl n gleich 1 ist. Nun ist aber $K - \{0\}$ nach (3.8) in einem homogenen Raum dicht, der nur noch zwei weitere Punkte umfaßt. Aus (3.11) folgt dann diese Behauptung.

(3.13). *Der Punkt x eines einfach zusammenhängenden Raumes E habe in E die Eigenschaft $P(x)$, wenn der Raum $E - \{x\}$ einfach zusammenhängend ist.*

(3.14). *Ein einfach zusammenhängender Raum E hat die Eigenschaft $P(x)$, wenn eine Umgebung U von x einfach zusammenhängend ist und die Eigenschaft $P(x)$ hat.*

Beweis. $E - \{x\}$ habe einen zusammenhängenden Überlagerungsraum F mit der Überlagerungsabbildung h. Der Teilraum $U - \{x\}$ wird von $h^{-1}(U - \{x\})$ schlicht überlagert. Da $U - \{x\}$ einfach zusammenhängend ist, zerfällt $h^{-1}(U - \{x\})$ in lauter zu $U - \{x\}$ homöomorphe Komponenten W_i (wo i eine Indexmenge I durchläuft). Die Umkehrabbildung der Einschränkung von h auf W_i heiße g_i. Zu jedem W_i kann man einen neuen Punkt x_i derart hinzufügen, daß g_i zu einem Homöomorphismus $\bar{g}_i$ von U auf $W_i \cup \{x_i\}$ so fortgesetzt

wird, daß $x_i = \bar{g}_i(x)$ gilt. Dann läßt sich h zu einer stetigen Abbildung $\bar{h}$ von $F \cup \bigcup_{i \in I} x_i$ auf E fortsetzen, indem in der Umgebung $W_i \cup \{x_i\}$ von x_i jeweils $\bar{h} = \bar{g}_i^{-1}$ gesetzt wird. Dann ist aber E von dem zusammenhängenden Raum $F \cup \bigcup_{i \in I} x_i$ schlicht überlagert bei der Überlagerungsabbildung $\bar{h}$. Da E einfach zusammenhängend ist, ist $\bar{h}$ ein Homöomorphismus. Dann enthält I nur ein Element und F ist homöomorph zu $E - \{x\}$. Also muß $E - \{x\}$ einfach zusammenhängend sein.

(3.15). *Jede Mannigfaltigkeit der Dimension n mit $n \geqq 3$ hat die Eigenschaft $P(x)$ für alle x.*

Beweis. Entfernt man aus einer wenigstens dreidimensionalen Kugel den Mittelpunkt, so ist der Rest immer noch einfach zusammenhängend.

(3.16). *G sei ein projektiver Limes von abzählbar vielen kompakten Liegruppen H_i, wo i die Menge N der natürlichen Zahlen durchläuft. Ist $i < j$, so bezeichne h_{ij} den Homomorphismus von H_j auf H_i; h_i sei der Homomorphismus von G auf H_i. (Vgl. z. B. [13] S. 23ff. oder [7] S. 176ff.) Z sei eine n-Zelle um 1 in H_1; da keine Verwechslungen zu befürchten sind, bezeichnen wir die neutralen Elemente aller auftretenden Gruppen stets mit 1; o. B. d. A. sei $n \geqq 3$. Die Umgebung $h_1^{-1}(Z)$ von 1 in G bezeichnen wir mit U, diese Umgebung können wir uns beliebig klein denken. Mit u bezeichnen wir einen von 1 verschiedenen Punkt in U. Ferner sei f eine Nullhomotopie einer geschlossenen Kurve in $U - \{1\}$, die in U auf u zusammengezogen wird, d. h. f ist eine stetige Abbildung des Quadrats $[0, 1] \times [0, 1]$ in U mit $f(r, 1) = f(0, s) = f(1, s') = u$, $f(r, 0) \neq 1$ für alle $r \in [0,1]$. Es gibt dann eine stetige Abbildung F des Würfels $W = [0, 1] \times [0, 1] \times [0, 1]$ in U, welche folgende Bedingungen erfüllt:*

a) *$F(r, s, 0) = f(r, s)$.*

b) *Falls $r = 0$, oder $r = 1$, oder $s = 0$, oder $s = 1$ gilt $F(r, s, t) = F(r, s, t')$ für alle $t, t' \in [0, 1]$.*

c) *$F(r, s, 1) \neq 1$ für alle $(r, s) \in [0, 1] \times [0, 1]$.*

Das bedeutet, daß die gegebene Homotopie in U stetig so deformiert werden kann, daß sie schließlich den Punkt 1 nicht mehr überstreicht.

Beweis. Die gesuchte Funktion F konstruieren wir durch Induktion. Eine Teilmenge I der Menge N der natürlichen Zahlen nennen wir regulär, wenn es zu jedem $i \in I$ eine Abbildung F_i von W in $h_{1i}^{-1}(Z)$ gibt mit folgenden Eigenschaften:

a_i) $F_i(r, s, 0) = h_i(f(r, s))$.

b_i) Falls $r = 0$, oder $r = 1$, oder $s = 0$ oder $s = 1$ gilt $F_i(r, s, t) = F_i(r, s, t')$ für alle $t, t' \in [0, 1]$.

c_i) $F_i(r, s, 1) \neq 1 \in H_i$ für alle $(r, s) \in [0, 1] \times [0, 1]$.

d_{ji}) Für alle $j < i$ aus I gilt stets $F_j = h_{ji} \circ F_i$.

Es gibt nun einen ersten Index $i \in N$, für den $h_i(f(r, 0)) \neq 1$ ist für alle $r \in [0, 1]$, sonst könnte nicht $f(r, 0) \neq 1$ in G sein. Es sei nun ohne Beschränkung der Allgemeinheit $i = 1$. Wir nehmen zunächst einmal als bewiesen an, daß in Z die Homotopie $(r, s) \to h_1(f(r, s))$ von 1 weggeschoben werden kann, d. h. daß eine Abbildung F_1 von W in Z existiert, die den Bedingungen a_1), b_1), c_1) und

trivialerweise d_{11}) genügt. Dann ist die Menge $\{1\}$ eine reguläre Teilmenge von N. Die regulären Mengen sind in bezug auf die Enthaltenseinsrelation induktiv geordnet. Es gibt also darunter maximale; bei dieser Aussage in der abzählbaren Menge N haben wir natürlich vom Auswahlaxiom keinen Gebrauch gemacht. Es sei nun J eine maximale reguläre Menge. Dann gehört zu J jedes $i \in N$, zu dem es ein $j \in J$ mit $i < j$ gibt: man braucht nur $F_i = h_{ij} \circ F_j$ zu setzen, um zu sehen, daß man J um i erweitern könnte, ohne die Regularität zu stören. Ferner kann J nicht beschränkt sein: Wäre etwa j das größte Element von J, so berücksichtigen wir, daß H_{j+1} einen Faserraum darstellt über H_j als Basis und mit dem Kern von h_{jj+1} als Faser. Nach Satz 11.7 von [12] S. 54 können wir die Homotopie F_j aus $h_{1j}^{-1}(Z)$ in $h_{1j+1}^{-1}(Z)$ hochheben, und zwar in der Weise, daß eine stetige Abbildung von W in H_{j+1} existiert, die a_{j+1}), b_{j+1}), c_{j+1}) und d_{jj+1}) und dann auch d_{ij+1}) für alle $i < j$ erfüllt. Dann gehört aber auch $j+1$ zu J, denn J war maximal regulär. Dies ist aber ein Widerspruch zur Bestimmung von j. Also ist $J = N$. Die Abbildung $(r, s, t) \to (F_i(r, s, t))_{i=1,\ldots}$ von W in $h_1^{-1}(Z)$ ist die gesuchte Abbildung F.

Es fehlt nun noch der Nachweis, daß die oben verwendete Abbildung F_1 existiert. Wir tragen ihn im folgenden Hilfssatz nach:

(3.17). *f und g seien Nullhomotopien ein und derselben geschlossenen Kurve in einem euklidischen Raum R_n, d. h. f und g seien stetige Abbildungen des Quadrates* $[0, 1] \times [0, 1]$ *in R^n mit $f(r, 0) = g(r, 0)$, $f(0, s) = g(0, s) = f(1, s) = g(1, s) = f(r, 1) = g(r, 1) = x \in R^n$. Dann sind die beiden Abbildungen f und g homotop, d.h. es existiert eine stetige Abbildung F des Würfels* $[0,1] \times [0,1] \times [0,1]$ *in R^n mit $F(r, s, 0) = f(r, s)$ und $F(r, s, 1) = g(r, s)$ sowie $F(r,0,t) = F(r,0,t')$ und $F(0, s, t) = x$.*

Beweis. F ist die stetige Fortsetzung einer auf dem Würfelrand definierten stetigen Abbildung in den R^n auf das Würfelinnere; eine solche Fortsetzung ist immer möglich, denn die zweite Homotopiegruppe von R^n ist trivial.

Mit diesem Hilfssatz ist leicht festzustellen, daß die in (3.16) gesuchte Abbildung F_1 existiert: Z hat die Eigenschaft $P(1)$. Ein in $Z - \{1\}$ liegender geschlossener Weg, der in Z auf $h_1(u) \neq 1$ zusammengezogen werden kann, kann auch in $Z - \{1\}$ auf $h_1(u)$ zusammengezogen werden. Die beiden Nullhomotopien sind in Z nach (3.17) homotop. Dies ist aber die Behauptung.

(3.18). *K sei eine in 0 punktierbare zusammenhängende lokalkompakte topologische Halbgruppe nichtendlicher Dimension. Dann ist die Gruppe $K - \{0\}$ ein Produkt einer zur additiven Gruppe R aller reellen Zahlen isomorphen Einparametergruppe E und dem in $K - \{0\}$ normalen direkten Produkt von abzählbar vielen ebenfalls in $K - \{0\}$ normalen einfachen und einfach zusammenhängenden Liegruppen L_i ($i = 1, 2, \ldots$).*

Beweis. $K' = K - \{0\}$ ist als lokalkompakte und zusammenhängende Gruppe kurvenzusammenhängend. Nach (3.7) hat jeder Punkt eine abzählbare Basis für seine Umgebungen. Nach (3.12) ist K' das Produkt einer zu R isomorphen Einparametergruppe E und einem größten kompakten Normalteiler B. Dieser ist nun ein projektiver Limes von abzählbar vielen Liegruppen. Unter Berücksichtigung von (1.1) und (1.4) wird die Behauptung bewiesen sein, wenn

der einfache Zusammenhang von $K' = K - \{0\}$ gezeigt ist. Der Punkt 0 hat in K eine Umgebung U mit der Eigenschaft, daß sich jeder geschlossene Weg, der von einem beliebigen Punkt $u \neq 0$ in U ausgeht und in ihn zurückläuft, ohne 0 zu treffen, in $U - \{0\}$ auf u zusammenziehen läßt, sobald er sich nur in ganz U auf u zusammenziehen läßt: (3.16) hat zusammen mit der Homogenität von K in der Tat diese Behauptung zur Folge. Nun sei $(r, s) \to f(r, s)$ die Nullhomotopie einer in K' verlaufenden geschlossenen Kurve, die in K etwa auf das Element 1 zusammengezogen werde; es sei also f eine stetige Abbildung des Quadrates $[0, 1] \times [0, 1]$ in K mit $f(0, s) = f(1, s) = f(r, 1) = 1$ und $f(r, 0) \neq 0$ für alle r und s aus $[0, 1]$. Das Pild M dieses Quadrates ist ein kompakter Teilraum von K. Nach (3.7) gibt es ein Element z in $K' = K - \{0\}$ derart, daß zM in U liegt. Die Abbildung $(r, s) \to zf(r, s)$ ist die Nullhomotopie eines geschlossenen Weges in $U - \{0\}$, der in U auf $z1 = z$ zusammengezogen wird. Es gibt jetzt eine stetige Abbildung g des Quadrates $[0, 1] \times [0, 1]$ in U derart, daß $g(0, s) = g(1, s) = g(r, 1) = z$ und $g(r, 0) = f(r, 0)$ gilt und $g(r, s) \neq 0$ ist für alle $(r, s) \in [0, 1] \times [0, 1]$. Die Abbildung $(r, s) \to z^{-1}g(r, s)$ von $[0, 1] \times [0, 1]$ ist eine Nullhomotopie der Ausgangskurve in K', die jetzt nicht nur in K, sondern sogar in $K' = K - \{0\}$ auf 1 zusammengezogen wird, denn ganz $U - \{0\}$ wird bei $x \to z^{-1}x$ in K' abgebildet. Die ersten Homotopiegruppen von K und K' sind also isomorph, d. h. K' ist einfach zusammenhängend.

Wir wenden uns nun dem endlichdimensionalen Fall zu:

(3.19). *K sei eine zusammenhängende, lokal kompakte, in 0 punktierbare topologische Halbgruppe endlicher Dimension, welche aber nicht kleiner sei als drei. Dann ist $K' = K - \{0\}$ isomorph zu der multiplikativen Gruppe aller von Null verschiedenen Quaternionen.*

Beweis. Wegen (3.15), (3.12), (1.1) und (2.2) wird es genügen, nachzuweisen, daß der eindeutig bestimmte größte kompakte Normalteiler B von K' die Dimension drei hat. B ist als halbeinfache nichtabelsche Liegruppe mindestens dreidimensional. h sei ein Isomorphismus der additiven Gruppe R aller reellen Zahlen auf eine Einparametergruppe E in K', mit der $K' = EB$ und $E \cap B = \{1\}$ gilt. Da 0 kein Adhärenzpunkt eines kompakten Teilraums von G sein kann, ist 0 Adhärenzpunkt von genau einem der Teilräume $G_1 = h(R - (-\infty, 0))B$ oder $G_2 = h(R - (0, \infty))B$. Es sei nämlich U eine offene Kugelumgebung von 0, die B nicht schneidet. $U \cap G_1$ und $U \cap G_2$ seien aber beide nicht leer; $U \cap G_1$ und $U \cap G_2$ sind dann punktfremde offene Teilmannigfaltigkeiten von $U - \{0\}$, deren Vereinigung gleich $U - \{0\}$ ist; aber $U - \{0\}$ ist zusammenhängend, da U wenigstens 4-dimensional ist. Es sei also etwa $U \cap G_1 = \emptyset$. Dann ist $G_2 \cup \{0\}$ eine Mannigfaltigkeit mit dem Rand B. Da B normal in K' ist, wirkt B als m-dimensionale Transformationsgruppe auf der $m + 1$-dimensionalen Mannigfaltigkeit $G_2 \cup \{0\}$ so, daß das singuläre Transitivitätssystem $\{0\}$ bei B festbleibt. Dann ist B homöomorph einer 0-, 1-, oder 3-Sphäre, wobei hier nur die Dreisphäre in Frage kommt: [8] S. 451, [9] S. 120.

(3.20). *K sei eine zweidimensionale zusammenhängende lokalkompakte in 0 punktierbare topologische Halbgruppe. Dann ist $K' = K - \{0\}$ isomorph zur multiplikativen Gruppe aller von 0 verschiedenen komplexen Zahlen.*

Beweis. Eine zweidimensionale Liegruppe ist einer Ebene, einem Zylinder oder einem Torus homöomorph (s. z. B. [6] S. 114). Nach (3.12) ist hier allein der Zylinder zutreffend, woraus die behauptete Isomorphie folgt (l. c.).

(3.21). *K sei eine eindimensionale zusammenhängende lokalkompakte topologische in 0 punktierbare Halbgruppe. Dann ist $K' = K - \{0\}$ isomorph zur multiplikativen Gruppe aller von Null verschiedenen reellen Zahlen.*

Beweis. Als eindimensionale Mannigfaltigkeit ist K in jedem Falle einfach zusammenhängend (3.6). K ist also kein homöomorphes Bild eines Kreises und K' zerfällt daher in genau zwei Komponenten. Die Komponente E von 1 ist ein Normalteiler vom Index 2 und ist zur additiven Gruppe R aller reellen Zahlen isomorph, da auch E sicher keine geschlossene Jordankurve und nicht kompakt ist. Die Abbildung $x \to y^{-1}xy$ ist für alle $y \in K'$ ein Automorphismus von E von der Ordnung 2; denn y^2 ist in jedem Fall in E enthalten, da y^2 kongruent 1 modulo E ist. Jedes Element von E aber bewirkt infolge der Kommutativität von E den identischen Automorphismus. Nun hat E nur einen nichttrivialen Automorphismus der Ordnung 2, nämlich denjenigen, der x in x^{-1} überführt. Für $x \in E$ bildet die Abbildung $y \to y^{-1}xy$ die zusammenhängende Nebenklasse von E stetig ab in die diskrete Menge $\{x, x^{-1}\}$. Da das Bild zusammenhängend sein muß, ist diese Abbildung konstant auf der ganzen Nebenklasse. Es gilt also entweder $xy = yx$ oder für alle $x \in E$ und alle y aus der Nebenklasse $xy = yx^{-1}$. Es werde nun y in der Nebenklasse festgehalten. Strebt x in E gegen 0, so strebt $y^{-1}xy$ gegen 0. Hingegen konvergiert x^{-1} nicht gegen einen Punkt von K, sondern gegen den von 0 verschiedenen Intervallendpunkt von E. Dieser Widerspruch zeigt, daß stets $xy = yx$ gelten muß. Dann ist aber K' eine kommutative Gruppe; da für ein z aus der Nebenklasse stets ein $x \in E$ vorhanden ist, so daß $z^2 = x^2$, gibt es eine Vertretergruppe B aus zwei Elementen, nämlich die von zx^{-1} erzeugte Gruppe. K' ist also ein direktes Produkt aus E und B und ist dann isomorph zur multiplikativen Gruppe aller von Null verschiedenen reellen Zahlen.

Die Hilfssätze (3.5), (3.4), (3.19), (3.20) und (3.21) ergeben zusammen den Beweis von Satz I der Einleitung, die Hilfssätze (3.5), (3.4) und (3.18) den Beweis von Satz II.

Anhang zu § 3

(3.3a). *Ein topologisches Monoid ist ein topologischer Raum, auf dem überall eine stetige Verknüpfung $(x, y) \to xy$ erklärt ist. Ein topologisches Monoid K nennen wir in J punktierbar, wenn folgende Bedingungen erfüllt sind:*

a) *K ist homogen und nichtdiskret.*

b) *J ist eine total unzusammenhängende, nirgends dichte nichtleere Teilmenge von K derart, daß $K - J$ eine in K offene topologische Loop ist.*

Eine topologische Loop ist dabei ein topologisches Monoid, indem die Gleichungen $ax = ya = b$ eindeutig und stetig nach x und y auflösbar sind und indem ein Element 1 mit $1x = x1 = x$ für alle x vorhanden ist.

Ein assoziatives punktierbares topologisches Monoid ist eine punktierbare topologische Halbgruppe. Eine topologische Loop mit Kürzungsregel ist eine

topologische Loop, in der für alle x die Identität $x^{-1}(xa) = (ax)\,x^{-1} = a$ mit einem allein von x abhängenden Element x^{-1} gilt. Dann gilt der Satz

(3.5) a. *Ein zusammenhängendes topologisches in J punktierbares topologisches Monoid, bei dem in K — J die Kürzungsregel gilt, kann nur eine aus einem Punkt bestehende Ausnahmemenge besitzen.*

Die Sätze (3.6), (3.7), (3.8), (3.9) haben alle eine entsprechende Verallgemeinerung, in der statt „topologische Halbgruppe" stets „topologisches Monoid" einzusetzen ist.

§ 4. Topologische Doppelloops

(4.1). *Wir bezeichnen als Doppelloop jede algebraische Struktur* $(K, +, .)$ *mit zwei Verknüpfungen* $(x, y) \to x + y$ *und* $(x, y) \to xy$, *die wir Addition bzw. Multiplikation nennen, und die folgenden Bedingungen genügen:*

a) *K ist eine Loop bezüglich der Addition mit dem neutralen Element* 0.

b) $K - \{0\}$ *ist eine Loop bezüglich der Multiplikation mit neutralem Element* 1. *Diese Loop nennen wir* $(K', .)$ *oder auch kurz* K'.

c) $x0 = 0x = 0$ *für alle* $x \in K$.

Da $1 \neq 0$ *ist, hat K mindestens zwei Elemente. Die Addition heißt potenzassoziativ, wenn jedes Element in einer Gruppe liegt, und diassoziativ, wenn immer zwei Elemente in einer Gruppe liegen. Es sei darauf hingewiesen, daß diese Bezeichnungsweise von der in* [3, 4] *benützten abgeht und sich der in* [1] *auf S.* 87 *eingeführten anschließt. Ist auf der Menge K eine Topologie erklärt, bezüglich der alle Verknüpfungen stetig sind, und zwar so, daß auch die Gleichungen* $a + x = y + a = b$ *und falls* $a \neq 0$ *und* $b \neq 0$ *auch die Gleichungen* $ax = ya = b$ *stetig auflösbar sind, so nennen wir K eine topologische Doppelloop. Insbesondere sind dann* $(K, +)$ *und* K' *topologische Loops* [*vgl.* (3.3) a]. *Eine Doppelloop nennen wir assoziativ, wenn ihre multiplikative Loop* K' *assoziativ, also eine Gruppe ist.*

(4.2). *Ist K eine topologische Doppelloop, so ist sie hinsichtlich der Multiplikation ein in* 0 *punktierbares Monoid* (3.3) a. *Ist K eine assoziative Doppelloop, so ist K hinsichtlich der Multiplikation eine in* 0 *punktierbare Halbgruppe. Falls K eine topologische Doppelloop ist, in der bezüglich der Multiplikation auf* K' *die Kürzungsregel gilt, so braucht die Bedingung* (4.1) c *nicht gefordert zu werden; sie ist dann nach* (3.4) *und dem Anhang zu* § 3 *eine Folgerung aus den Bedingungen* (4.1 a), (4.1 b) *und den topologischen Voraussetzungen, denn* (4.1 a) *sichert die Homogenität.*

Daraus und aus Satz I folgt aber Satz I', während Satz II zusammen mit dem gerade bewiesenen Hilfssatz (4.2) den Satz II' nach sich zieht.

(4.3). *Wir bezeichnen mit* $\mathscr{K}$ *die Kategorie aller zusammenhängenden lokalkompakten endlichdimensionalen topologischen assoziativen Doppelloops, mit* $\mathscr{K}_p$ *die Teilkategorie aller Doppelloops aus* $\mathscr{K}$ *mit potenzassoziativer Addition, mit* $\mathscr{K}_d$ *die Teilkategorie von* $\mathscr{K}_p$, *bei deren Elementen die Addition sogar diassoziativ ist, und schließlich mit* $\mathscr{K}_a$ *diejenige Teilkategorie von* $\mathscr{K}_d$, *bei der sie assoziativ ist. Diejenige Teilkategorie einer Kategorie* $\mathscr{K}_m$ $(m = p, d, a)$, *in deren*

sämtlichen Elementen das Rechtsdistributivgesetz

a) $$(x+y)z = xz + yz$$

gilt, bezeichnen wir mit $\mathcal{K}'_m$; wird auch noch die Gültigkeit des Linksdistributivgesetzes

b) $$z(x+y) = zx + zy$$

gefordert, so wird dadurch eine Teilkategorie $\mathcal{K}''_m$ von $\mathcal{K}'_m$ ausgezeichnet, die wir die Kategorie der zusammenhängenden lokalkompakten topologischen Neokörper nennen. Die Kategorie $\mathcal{K}$ teilen wir in Klassen $\mathcal{D}(l)$ ein in der Weise, daß in $\mathcal{D}(l)$ gerade die l-dimensionalen Doppelloops liegen.

(4.4). *Die Kategorie $\mathcal{K}'_p \cap \mathcal{D}(4)$ zerfällt in Äquivalenzklassen $\mathcal{C}(\lambda)$ so, daß jeder reellen Zahl λ genau eine Klasse zugeordnet ist und zwei isomorphe Doppelloops immer in einer und derselben Klasse liegen.*

Beweis. Es sei $(K, +, .)$ eine Doppelloop aus der Kategorie $\mathcal{K}'_p$. Dann enthält K einen eindeutig bestimmten, zum Körper der reellen Zahlen isomorphen Unterkörper H: Die von 1 in K algebraisch erzeugte Unterdoppelloop ist isomorph zum Körper der rationalen Zahlen [4] (1.3). Die abgeschlossene Hülle dieses Unterkörpers von K ist ein lokalkompakter Unterkörper H von K. Nach [4] (1.10) und [4] (3.5) ist dieser Unterkörper nicht total unzusammenhängend und ist daher isomorph zum Körper der reellen Zahlen. (Es sei angemerkt, daß in der erwähnten Arbeit die Elemente der Unterdoppelloop bei Linksmultiplikation als Automorphismen der Addition wirken, während das jetzt für die Rechtsmultiplikationen der Fall ist.) Die Zusammenhangskomponente von 1 in $H - \{0\}$ ist eine eindeutig festgelegte Einparameteruntergruppe E von K'. Sie bestimmt eindeutig eine eindimensionale Unteralgebra der Liealgebra von K' mit einer erzeugenden infinitesimalen Transformation $\mathfrak{l}'_0$; diese sei so normiert, daß bei isomorpher Einbettung des Körpers der reellen Zahlen in K mit geeigneten kanonischen Koordinaten erster Art die Beziehung $\exp t = e^t$ gilt. Dann läßt sich $\mathfrak{l}'_0$ gemäß (2.1) zu einer Basis $\mathfrak{l}'_0, \mathfrak{l}_1, \mathfrak{l}_2, \mathfrak{l}_3$ der Lie-Algebra ergänzen, die folgende Multiplikationstafel hat:

	$\mathfrak{l}'_0$	$\mathfrak{l}_1$	$\mathfrak{l}_2$	$\mathfrak{l}_3$
$\mathfrak{l}'_0$	0	0	$\lambda\mathfrak{l}_3$	$-\lambda\mathfrak{l}_2$
$\mathfrak{l}_1$	0	0	$\mathfrak{l}_3$	$-\mathfrak{l}_2$
$\mathfrak{l}_2$	$-\lambda\mathfrak{l}_3$	$-\mathfrak{l}_3$	0	$\mathfrak{l}_1$
$\mathfrak{l}_3$	$\lambda\mathfrak{l}_2$	$\mathfrak{l}_2$	$-\mathfrak{l}_1$	0

Dabei ist die reelle Zahl λ durch die Vorgabe der Einparametergruppe E dem Betrag nach eindeutig festgelegt; ein Basiswechsel, bei dem $\mathfrak{l}'_0$ in $-\mathfrak{l}'_0$ übergeht, ruft eine Multiplikationstafel hervor, in der λ gegenüber der obenstehenden durch $-\lambda$ ersetzt ist. Zu beiden Parameterwerten λ und $-\lambda$ gehören natürlich isomorphe Multiplikationsloops, aber keine isomorphen Doppelloops. Es seien nämlich $(K_1, +, .)$ und $(K_2, +, .)$ zwei isomorphe Doppelloops und φ ein Isomorphismus von K_1 auf K_2; zu K_1 sei nach dem in dieser Nummer dargelegten Verfahren der Parameter λ_1 ermittelt und zu K_2 der Parameter λ_2.

Denken wir uns sowohl in K_1 wie in K_2 den Körper der reellen Zahlen eingebettet, was auf eine und nur eine Weise möglich ist, so gilt $\varphi(r) = r$ für alle r aus dem Körper der reellen Zahlen. Zu φ existiert ein Isomorphismus $d\varphi$ von der zu K_1' gehörigen Liealgebra $\mathfrak{G}_1$ auf die zu K_2' gehörige Liealgebra $\mathfrak{G}_2$, für den $\varphi(\exp \mathfrak{x}) = \exp d\varphi(\mathfrak{x})$ für jede infinitesimale Transformation $\mathfrak{x}$ aus $\mathfrak{G}_1$ gilt. Ist $\mathfrak{l}_{0m}', \mathfrak{l}_{1m}, \mathfrak{l}_{2m}, \mathfrak{l}_{3m}$ eine nach dem obenstehenden Vorgehen gewonnene Basis von $\mathfrak{G}_m$ $(m = 1, 2)$, so gilt wegen $\varphi(r) = r$ für alle reellen r auf jeden Fall die Gleichung $d\varphi(\mathfrak{l}_{01}') = \mathfrak{l}_{02}'$. Ist nun $\lambda_1 = 0$, so ist der reelle Unterkörper von K_1 zentral in dem Sinne, daß seine Multiplikationsgruppe im Zentrum von K_1' liegt. Da das Zentrum von K_1' bei φ in das Zentrum von K_2' übergeht, ist auch der reelle Unterkörper von K_2 zentral; die von $\mathfrak{l}_{02}'$ erzeugte Unteralgebra von $\mathfrak{G}_2$ ist dann das Zentrum von $\mathfrak{G}_2$ und es folgt $\lambda_2 = 0$. Ist aber $\lambda_1 \neq 0$, so gilt $d\varphi(\mathfrak{l}_{11} - \lambda_1^{-1}\mathfrak{l}_{01}') = d\varphi\mathfrak{l}_{11} - \lambda_1^{-1}\mathfrak{l}_{02}'$ und $d\varphi\mathfrak{l}_{11} - \lambda_1^{-1}\mathfrak{l}_{02}'$ erzeugt das Zentrum von $\mathfrak{G}_2$. Dann ist λ_2 von 0 verschieden; wenn aber λ_1 und λ_2 von 0 verschieden sind, dann ist $\mathfrak{l}_{1m}$ aus $\mathfrak{l}_{0m}'$ für $m = 1, 2$ eindeutig bestimmt und es muß $d\varphi(\mathfrak{l}_{11}) = \mathfrak{l}_{12}$ gelten; dann ist auch $\lambda_1 = \lambda_2$.

(4.5). $(K, +, .)$ *sei aus der Kategorie* $\mathcal{K}_p'' \cap \mathcal{D}(4)$ *(vgl.* (3.5)*). Wir nennen K dann auch einen quaternalen Neokörper mit potenzassoziativer Addition. Dann gibt es auf dem Raum K zu jedem reellen* λ *eine neue Multiplikation* $(x, y) \to x \circ y$ *derart, daß* $(K, +, \circ)$ *der Klasse* $\mathcal{C}(\lambda)$ *der Kategorie* $\mathcal{K}_p' \cap \mathcal{D}(4)$ *angehört. Die Klasse* $\mathcal{C}(0)$ *enthält* $\mathcal{K}_p'' \cap \mathcal{D}(4)$.

Beweis. $(x, y) \to x \circ y$ sei die in (2.3) eingeführte Verknüpfung, wobei wir uns in der Liealgebra von K' auf eine Basis $\mathfrak{l}_0, \mathfrak{l}_1, \mathfrak{l}_2, \mathfrak{l}_3$ stützen, wie sie in (2.1) angegeben war. Wegen der Gültigkeit beider Distributivgesetze haben wir die Identität $(x + y)^s = x^s + y^s$, wenn u^s wie in (2.3) erklärt wird. Es gilt das Rechtsdistributivgesetz für die $\circ$-Multiplikation: $(x + y) \circ z = (x + y)^s z = (x^s + y^s) z = x^s z + y^s z = x \circ z + y \circ z$. Das Linksdistributivgesetz gilt nicht. Nach der Definition des Homomorphismus h in (2.3) gibt es in dem reellen Unterkörper von K ein Element r derart, daß $h(r)$ nicht im Zentrum von K' liegt, da dieses die von $\mathfrak{l}_1$ erzeugte Untergruppe nicht enthält. Wir wählen r insbesondere so, daß $h(r + r) = -1$ ist, wobei -1 in dem reellen Unterkörper von K liegt und das einzige Element darstellt, das in der Gruppe K' die Ordnung 2 hat. Dann gilt $x \circ (r + r) = (-1)^{-1} x (-1)\,(r + r) = 2xr$, während $x \circ r + x \circ r = 2x \circ r$ davon immer verschieden ist, wenn x nicht aus der von $\mathfrak{l}_0$ und $\mathfrak{l}_1$ erzeugten Untergruppe stammt.

(4.6). $(K, +, .)$ *sei aus der Kategorie* $\mathcal{K}_a' \cap \mathcal{D}(4)$. *Dann ist die Addition kommutativ. Liegt K in der Klasse* $\mathcal{C}(\lambda)$*, und ist* λ *von 0 verschieden, so enthält K einen und nur einen Unterkörper, der isomorph zum Körper der komplexen Zahlen ist. Ist* $\lambda = 0$*, dann liegt K in* $\mathcal{K}_a''$*, d. h. K ist dann isomorph zum Schiefkörper der Quaternionen.*

Beweis. $(K, +, .)$ enthält einen wohlbestimmten Unterkörper isomorph zum Körper der reellen Zahlen. Ist wieder $\mathfrak{l}_0', \mathfrak{l}_1, \mathfrak{l}_2, \mathfrak{l}_3$ eine Basis der Lie-Algebra von K' gemäß (4.4), so erzeugen die infinitesimalen Transformationen $\mathfrak{l}_0'$ und $\mathfrak{l}_1$ eine Untergruppe, die zur multiplikativen Gruppe der von 0 verschiedenen

Zahlen isomorph ist. Falls $\lambda \neq 0$, ist dies die einzige Gruppe dieser Art, da das Lieprodukt von $\mathfrak{l}_0'$ mit irgendeiner nicht in der von $\mathfrak{l}_0'$ und $\mathfrak{l}_1$ erzeugten Unteralgebra liegenden infinitesimalen Transformation nicht 0 ist. Zu der angegebenen Gruppe existiert in K nach [4] (4.10) ein Unterkörper, der zum Körper der komplexen Zahlen isomorph ist, und dessen Multiplikationsgruppe gerade mit ihr übereinstimmt. Ist aber $\lambda = 0$, so gehört nach [4] (4.10) zu jeder Torusuntergruppe von K' ein und nur ein Unterkörper, der zum Körper der komplexen Zahlen isomorph ist. Dann ist nach [4] (4.12) K isomorph zum Schiefkörper der Quaternionen. Es sei darauf hingewiesen, daß an den aus [4] zitierten Stellen (4.10) und (4.12) beide Distributivgesetze gefordert sind. In den Beweisen zu beiden Sätzen gehen aber nur kommutative Multiplikationsgruppen ein, nämlich solche von komplexen Unterkörpern, die im Fall $\lambda = 0$ in reicher Anzahl vorhanden sind; liegt doch jeder beliebige Punkt in einer von $\mathfrak{l}_0 = \mathfrak{l}_0'$ und irgendeiner von $\mathfrak{l}_0$ linear unabhängigen linearen Transformation erzeugten, zur komplexen Multiplikationsgruppe isomorphen Untergruppe. Die Kommutativität der Addition folgt im allgemeinen Fall aus [4] (4.11).

Ist die Addition assoziativ, so wirken die Elemente von K als lineare Abbildungen auf dem durch die Additionsgruppe gegebenen vierdimensionalen Vektorraum über dem Körper der reellen Zahlen. Die Struktur von K ist völlig festgelegt, wenn von jedem Element aus K festliegt, wohin es das Element 1 abbildet. Da die Multiplikationsgruppe von K bei festem, von 0 verschiedenem λ völlig eindeutig festliegt, und zwar dadurch, daß die infinitesimalen Transformationen $\mathfrak{l}_0$ und $\mathfrak{l}_1$ eindeutig bestimmt sind, sind alle K, die zu ein und demselben Parameterwert λ gehören, isomorph (vgl. [5]).

Unsere Ergebnisse über die zusammenhängenden, lokalkompakten topologischen Doppelloops mit assoziativer Multiplikation fassen wir in dem folgenden Satz zusammen; bei seiner Formulierung vereinbaren wir, daß alle isomorphen Doppelloops in $\mathcal{K}$ identifiziert werden, so seien also in Wirklichkeit die Elemente von $\mathcal{K}$ Klassen isomorpher Doppelloops.

Satz III.

(4.7). a) $\mathcal{D}(l) = \emptyset$, *falls* $l \neq 1, 2,$ *oder* 4.

b) $\mathcal{K}_p' \cap \mathcal{D}(1)$ *enthält nur den Körper der reellen Zahlen.*

c) $\mathcal{K}_p' \cap \mathcal{D}(2)$ *enthält nur Neokörper mit einer zur multiplikativen Halbgruppe der komplexen Zahlen isomorphen Multiplikationshalbgruppe und einem reellen Unterkörper.*

d) $\mathcal{K}_d' \cap \mathcal{D}(2)$ *enthält nur den Körper der komplexen Zahlen.*

e) $\mathcal{K}_p'' \cap \mathcal{D}(2) - \mathcal{K}_d'' \cap \mathcal{D}(2)$ *ist nichtleer.*

f) $\mathcal{K}_p' \cap \mathcal{D}(4) \cap \mathcal{C}(\lambda)$ *ist nichtleer für alle reellen* λ. *Ist* $\lambda \neq 0$, *so gilt im allgemeinen nur ein Distributivgesetz. Alle Doppelloops aus* $\mathcal{K}_p' \cap \mathcal{D}(4) \cap \mathcal{C}(\lambda)$ *enthalten einen reellen Unterkörper.*

g) *Alle Doppelloops aus* $\mathcal{K}_d' \cap \mathcal{D}(4) \cap \mathcal{C}(\lambda)$ *enthalten einen und im Fall* $\lambda \neq 0$ *nur einen komplexen Unterkörper.*

h) $\mathcal{K}_d' \cap \mathcal{D}(4) \cap \mathcal{C}(0)$ *enthält nur den Schiefkörper der Quaternionen.*

Beweis.

a) Satz I. b) Satz I, [3] (2.4). c) Satz I, vgl. auch [4] § 3.
d) [3] S. 227. e) [3] (4.10). f) (4.5).
g) (4.6). h) (4.6).

Literatur

[1] Bruck, R. H.: A survey on binary systems. Erg. Math. 20. Heft (1958). — [2] Chevalley, C.: Theory of Lie groups I. Princeton Univ. Press 1946. — [3] Hofmann, K. H.: Topologische Doppelloops. Math. Z. 70, 213—230 (1958). — [4] Hofmann, K. H.: Topologische distributive Doppelloops. Math. Z. 71, 36—68 (1959). — [5] Kalscheuer, F.: Die Bestimmung aller stetigen Fastkörper. Abh. math. Sem. Univ. Hamburg 13, 413—435 (1940). — [6] Kerékjártó, B. v.: Geometrische Theorie der zweigliedrigen kontinuierlichen Gruppen. Abh. math. Sem. Univ. Hamburg 8, 107—114 (1931). — [7] Montgomery, D., and L. Zippin: Topological transformation groups. Intersci. Tracts in pure appl. Math. 1. Heft (1955). — [8] Mostert, P. S.: On a compact Lie group acting on a manifold. Ann. of Math. 65, 447—455 (1957). — [9] Mostert, P. S., and A. L. Shields: On the structure of semigroups on a compact manifold with boundary. Ann. of Math. 5, 117—143 (1957). — [10] Novosad, R. S.: Simply connected spaces. Trans. Amer. math. Soc. 79, 216—228 (1955). — [11] Séminaire „S. Lie“: Théorie des algèbres de Lie, Topologie des groupes de Lie. 1954/1955, Ec. Norm. Sup., Secrét. math. — [12] Steenrod, N.: The topology of fibre bundles. Princeton Univ. Press 1951. — [13] Weil, A.: L'intégration dans les groupes topologiques et ses applications. Act. scient. et ind. 8, 69 (1940)—1145 (1951).

(Eingegangen am 25. März 1959)

Schaefer, H.
Math. Annalen 138, 259—286 (1959)

Halbgeordnete lokalkonvexe Vektorräume. II

Von

Helmut Schaefer in Pullman, Wash.

Einleitung

Die vorliegende Arbeit ist eine unmittelbare Fortsetzung von [13′] (im folgenden mit HV zitiert)[1]). Die Abschnitte sind im Anschluß an HV fortlaufend numeriert. Mit Ausnahme der Abschnitte 6 und 7, die Ergänzungen zu HV und neue Ergebnisse im Rahmen von HV bringen, ist die gegenwärtige Untersuchung dem Studium stetiger linearer Abbildungen eines halbgeordneten lokalkonvexen Raumes E in einen zweiten F gewidmet. Die zueinander dualen Begriffe des normalen und des $\mathfrak{S}$-Kegels, vgl. (1.5), erweisen sich auch hier wieder als fundamental. Die Diskussion variiert zwischen der Betrachtung einer einzelnen positiven linearen Abbildung (d. h. eines Ordnungshomomorphismus, HV p. 119) und dem Studium algebraischer und topologischer Eigenschaften des Kegels $\mathfrak{K}$ aller Ordnungshomomorphismen im Raum $\mathfrak{L}(E,F)$ stetiger linearer Abbildungen, der mit einer Reihe von (lokalkonvexen) sog. $\mathfrak{S}$-Topologien versehen werden kann (s. Abschnitt 8). Beide Aspekte sind naturgemäß eng miteinander verknüpft (vgl. als Beispiel hierzu (11.1) und das zugehörige Corollar). Die benutzten Hilfsmittel sind (mit Ausnahme des in Abschnitt 8 verwendeten Tensorproduktes) dieselben wir in HV (HV, p. 118 bis 120). Wir geben eine kurze Übersicht über den Inhalt der einzelnen Abschnitte.

Abschnitt 6 beschäftigt sich mit der Übertragung der wichtigsten Ergebnisse von Abschnitt 1 in HV auf den Fall linearer Räume mit komplexem Skalarkörper. (Durchgehend bezeichnen wir mit $\mathbf{N}$ [bzw. $\mathbf{R}$, bzw. $\mathbf{C}$] die Mengen der natürlichen [bzw. reellen, bzw. komplexen] Zahlen.) Mit einer geeigneten Definition (Def. 6) eines normalen bzw. $\mathfrak{S}$-Kegels in einem komplexen lokalkonvexen Raum geht die Erweiterung auf den komplexen Fall reibungslos vor sich. Hierbei ist der Begriff der positiven Linearform neu zu fassen (Def. 7). Die Methode der Verallgemeinerung ist genügend auseinandergesetzt, um auch die Übertragbarkeit der meisten Resultate der Abschnitte 2 bis 5 anzuzeigen. Der Einschluß des komplexen Falles ist nicht nur wegen der größeren Allgemeinheit wünschenswert, sondern auch für Spektralbetrachtungen äußerst nützlich (Abschnitte 10 und 11).

[1]) Um eine weitgehende Wiederholung des Literaturverzeichnisses von HV zu vermeiden, zitieren wir in HV aufgeführte Arbeiten mit ihren dortigen Nummern. Die hinzukommenden Literaturhinweise sind neu durchnumeriert, und jede Nummer ist mit einem Akzent ′ versehen. Sätze und Definitionen aus HV werden unter ihrer dortigen Nummer zitiert. Allgemein ist $(u.v)$ Satz v in Abschnitt u.

Von Abschnitt 7 an werden die betrachteten Räume im allgemeinen als komplex vorausgesetzt. Die Hauptresultate von Abschnitt 7 sind die Konvergenzsätze (7.2) und (7.3). Ist H eine gerichtete Menge in $E[\mathfrak{T}]$, $E[\mathfrak{T}]$ halbgeordnet mit normalem positiven Kegel K, so folgt aus der schwachen Konvergenz des Filters $\mathfrak{F}(H)$ der Enden von H die Konvergenz von $\mathfrak{F}$ für $\mathfrak{T}$. Der Satz kann verschärft werden: Ist E halbreflexiv, K normal, H (aufsteigend) gerichtet und (nach oben ordnungs-) beschränkt, so konvergiert $\mathfrak{F}$ bereits unter diesen Voraussetzungen (7.3). Dieser Satz kann als Verallgemeinerung der Aussage gedeutet werden, daß eine beschränkte monotone Folge reeller Zahlen konvergiert. (7.4) zeigt, daß der zu einem echten Kegel mit inneren Punkten konjugierte Kegel eine schwach kompakte Basis besitzt, woraus ein Konvergenzsatz für positive Filter folgt (7.5). Normale Kegel mit inneren Punkten sind selten (7.6).

Der zweite Teil der Untersuchung beginnt mit Abschnitt 8, wo einige Beziehungen zwischen den positiven Kegeln K, H in E, F und dem „positiven" Kegel $\mathfrak{K} \subset \mathfrak{L}(E, F)$ aufgestellt werden. $\mathfrak{K}$ ist unter einfachen Bedingungen normal für die $\mathfrak{S}$-Topologie. Die Frage nach der BZ-Eigenschaft von $\mathfrak{K}$ und die damit zusammenhängende Frage, wann $\mathfrak{L} = \mathfrak{K} - \mathfrak{K}$ ist, ist wesentlich schwieriger zu beantworten. Die einfache notwendige Bedingung (8.4) scheint bei weitem nicht hinreichend zu sein; sie ist es jedenfalls, wenn E nuklear (und tonneliert) und F ein B-Raum ist [Cor. zu (8.5)]. Allgemeiner ist jede nukleare Abbildung unter vernünftigen Bedingungen zerlegbar (8.5). Dies schließt die Abbildungen endlichen Ranges ein, für die sich alles aus (1.3) bzw. (6.1) ablesen läßt.

Abschnitt 9 gibt in (9.1) bis (9.3) Voraussetzungen, unter denen jede positive lineare Abbildung stetig ist. Daß hierbei die Kegel mit inneren Punkten (9.1) und die BZ-Kegel in bornologischen Räumen (9.3) hervortreten würden, war nach einem bekannten Satz über Linearformen ([7], p. 75) und (2.8) zu erwarten. Bei diesen Sätzen ist jedesmal wesentlich, daß der Kegel H im Bildraum F normal ist. [Daß positive lineare Abbildungen für die Ordnungstopologien, vgl. (4.4), auf E und F stetig sind, ist beinahe selbstverständlich[2].] (7.3) reicht noch bis zu einem Konvergenzsatz für den Filter der Enden einer (für die durch $\mathfrak{K}$ in $\mathfrak{L}$ erzeugte Ordnungsstruktur) gerichteten Menge stetiger linearer Abbildungen (9.4), der für Endomorphismen reflexiver Räume eine besonders einfache Form annimmt.

Der im 10. Abschnitt diskutierte Problemkreis ist, wenn auch im allgemeinen nur für normierte Räume (oder unter noch engeren Bedingungen, und mit reicheren Ergebnissen, [3']), Gegenstand einer Reihe von Untersuchungen gewesen. Der erste allgemeine Satz dürfte in [17] zu finden sein, Verschärfungen und Erweiterungen in [2] bis [4]. Der erste für beliebige lokalkonvexe Räume gültige Satz über die Existenz eines positiven Eigenwertes mit positivem Eigenvektor (für eine gegebene Halbordnung) eines positiven, kompakten Endomorphismus findet sich in [20]. Eine weitere Verallgemeinerung für

[2]) und übrigens nach (4.4) eine Folgerung aus (9.1).

normierte Räume, die die Stetigkeit und Kompaktheit von T nur innerhalb des positiven Kegels voraussetzt, hat in [4'] eine sehr ausführliche Darstellung gefunden. Allerdings enthält auch [20] einen Satz, der nur auf die Verhältnisse innerhalb des positiven Kegels Bezug nimmt[3]). Das Hauptergebnis (10.4) in diesem Abschnitt fällt, wenn E ein normierter Raum ist, mit dem Hauptresultat von [4'] zusammen. Zahlreiche der in [4'] bewiesenen Aussagen [z. B. Lemma (3.1), (3.2)] sind nichts anderes als die entsprechenden bekannten Sätze für die vor (10.1) eingeführte Topologie $\mathfrak{T}_0$ auf der linearen Hülle E_K von K [vgl. hierzu (10.1) bis (10.3)]. Um zu (10.4) für beliebige lokalkonvexe Räume zu gelangen, bedarf es einer gegenüber [4'] wesentlich verfeinerten Technik, da Kategorie-Argumente, beschränkte Nullumgebungen, Auswahl konvergenter Teilfolgen nicht mehr zur Verfügung stehen. Die volle Last des Beweises ruht auf der K-Kompaktheit von T (Def. 8); auch Vollständigkeitseigenschaften von K sind entbehrlich. Um die Darstellung möglichst unabhängig zu gestalten, wird eine Reihe von Spektralaussagen in Form von Hilfssätzen (Lemma 1 bis 3) aufgestellt, deren Beweis (auf eine, soweit es Verf. bekannt ist, neuartige Weise) zum Teil nur skizziert wird. (10.5) ist eine befriedigende Verallgemeinerung des Krein-Rutman-Theorems [17]. Es sei noch bemerkt, daß in diesem Abschnitt der Begriff des normalen Kegels in den Hintergrund tritt; die Echtheit (und Abgeschlossenheit) von K ist bei K-kompaktem T für alle Schlüsse ausreichend [vgl. jedoch (10.7)].

Unter Beschränkung (im wesentlichen) auf normierte und Hilbertsche Räume bringt der 11. Abschnitt für größere Operatorenklassen Beiträge über den Zusammenhang zwischen dem Vorhandensein reeller Zahlen ≥ 0 im Spektrum von T und der Existenz unter T invarianter Kegel. (11.1) zeigt, daß die Normalität eines Kegels $\mathfrak{K}$ in einer *B-Algebra mit Einselement ausreicht, um die Zugehörigkeit des Spektralradius $r(a)$ zum Spektrum $\sigma(a)$ für jedes $a \in \mathfrak{K}$ zu sichern. Das zugehörige Corollar, das (11.1) auf Operatorenalgebren in B-Räumen anwendet, geht auf [4] zurück. (11.2) zeigt einen weiteren Zusammenhang zwischen der Existenz T-invarianter Kegel und dem Spektrum von T. Schließlich wird in (11.3) ein Zusammenhang zwischen positiven symmetrischen Operatoren (d. h. stetigen Endomorphismen, die symmetrisch bzw. hermetisch sind und ein nichtnegatives Spektrum besitzen) und sehr speziellen Typen konvexer Kegel in einem Hilbertraum hergestellt.

6. Komplexe Räume

Wir erweitern in diesem Abschnitt die wichtigsten der in HV, Abschn. 1 und 2, entwickelten Resultate über normale und $\mathfrak{S}$-Kegel auf lokalkonvexe lineare Räume über dem Körper $\mathbf{C}$ der komplexen Zahlen. Hierzu erinnern wir zunächst an folgenden Sachverhalt (vgl. [7], p. 106). Es sei L (bzw. $L[\mathfrak{T}]$) ein komplexer, linearer (bzw. topologischer linearer) Raum. Durch Einschränkung des Skalarkörpers auf $\mathbf{R}$ erhält man einen reellen Raum L_0 (bzw. $L_0[\mathfrak{T}]$), der

---

[3]) Satz 3.1. Eine ausführlichere Diskussion des nichtlinearen Falles wird in einer demnächst im Pacific Journal of Math. erscheinenden Arbeit gegeben.

als abelsche Gruppe (und als topologischer Raum) mit L übereinstimmt. Wir nennen L_0 den zu L *assoziierten* reellen Raum. L_0, mit der topologischen Struktur von L, ist ein topologischer linearer Raum, d. h. Topologie und algebraische Struktur sind miteinander verträglich.

Ist umgekehrt ein reeller (topologischer) linearer Raum E gegeben, so ist E assoziiert zu einem komplexen (topologischen) linearen Raum F, d. h. $E = F_0$, genau wenn in E ein (topologischer) Automorphismus φ mit $\varphi^2 = -1$ existiert. Ist $E = F_0$ für ein F, so ist offenbar $x \to i x$ ein Automorphismus der gewünschten Art. Ist umgekehrt φ ein solcher Automorphismus von E, so ist die zu E assoziierte (topologische) abelsche Gruppe, versehen mit der Skalarmultiplikation $(\lambda, x) \to \alpha x + \beta \varphi(x)$ $(\lambda = \alpha + i\beta; \alpha, \beta \in \mathbf{R})$, ein komplexer (topologischer) linearer Raum F mit $E = F_0$.

Ein komplexer topologischer linearer Raum L ist lokalkonvex [7], wenn der assoziierte reelle Raum L_0 lokalkonvex ist. Entsprechend definieren wir:

Ein *komplexer* linearer Raum L ist *halbgeordnet* (bzw. *regulär halbgeordnet*), wenn sein assoziierter reeller Raum L_0 halbgeordnet (bzw. regulär halbgeordnet) ist (HV, p. 118 u. 126).

Die Menge der für eine solche Halbordnung positiven Elemente bildet wieder einen konvexen Kegel mit dem Scheitel 0[4]) in L. Zur Definition von normalen bzw. $\mathfrak{S}$-Kegeln (HV, Def. 1 u. 2) verabreden wir:

Definition 6. *Ein Kegel K in einem komplexen lokalkonvexen Raum E heiße normal (bzw. $\mathfrak{S}$-Kegel), wenn K normal (bzw. $\mathfrak{S}$-Kegel) in dem zu E assoziierten reellen Raum E_0 ist.*

Zur Definition des konjugierten Kegels im Fall eines komplexen Dualsystems setzen wir fest:

Definition 7. *Es sei $\langle F, G\rangle$ ein komplexes Dualsystem, K ein Kegel in F, K' (bzw. H') der größte Kegel in G, so daß die kanonische Bilinearform auf $K \times K'$ einen nichtnegativen Realteil (bzw. auf $K \times H'$ einen nichtnegativen Real- und Imaginärteil) besitzt. Die Elemente von K' (bzw. H') heißen schwach positiv (bzw. stark positiv) auf K.*

Offenbar ist K' die zu $-K$ polare Menge in G, und $H' = K' \cap (-iK)'$. Das Zerlegungstheorem (1.3) nimmt jetzt die Form an

(6.1). *Ist K ein normaler Kegel in $E[\mathfrak{T}]$, so gilt $E' = K' - K'$, d. h. jede stetige Linearform auf E ist Differenz zweier stetiger, schwach positiver Linearformen.*

Beweis. Jede auf E stetige Linearform schreibt sich $x \to h(x) = f(x) - i f(i x)$, wo f eine reelle stetige Linearform auf E (d. h. eine stetige Linearform auf E_0) ist. Da K in E_0 normal ist, gilt nach (1.3) $f = f_1 - f_2$, und f_j ist der Realteil einer auf K schwach positiven, auf E stetigen Linearform $h_j (j = 1, 2)$. Offenbar ist $h = h_1 - h_2$.

Für die Zerlegbarkeit von E' nach H' gilt

(6.2). *Es sei K ein Kegel in E, so daß $K + iK$ normal ist. Jede auf E stetige Linearform ist Differenz zweier auf K stark positiver stetiger Linearformen, d. h. $E' = H' - H'$.*

[4]) Wie in HV bezeichnen wir einen solchen stets als *Kegel* (schlechthin).

Beweis. Da $x \to -ix$ ein topologischer Automorphismus von E_0 ist, ist mit $K + iK$ auch $K - iK$ normal. Nach (6.1) gestattet jede auf E stetige Linearform h eine Darstellung $h = h_1 - h_2$, wo h_1 und h_2 stetig und schwach positiv auf $K - iK$ sind. Nun gilt $h_1(x) = f(x) + if(-ix)$, d. h. h_1 ist stark positiv auf K; entsprechendes gilt für h_2.

Die Erweiterung von (1.6) auf komplexe Räume ist

(6.3). *Sei E ein komplexer lokalkonvexer Raum, K ein Kegel in E. Damit $E' = K' - K'$ (bzw. $E' = H' - H'$) gelte, ist notwendig und hinreichend, daß K (bzw. $K + iK$) normal sei für die schwache Topologie $\sigma(E, E')$.*

Beweis. Wir haben nur die Notwendigkeit der Bedingung zu zeigen. Wenn $E' = K' - K'$ gilt, so erzeugen die Halbnormen ($\langle x, x'\rangle$ der Wert der kanonischen Bilinearform)

$$(*) \qquad x \to |\mathsf{R}\langle x, x'\rangle| \qquad (x' \in K')$$

die schwache Topologie auf E ($\mathsf{R}\langle x, x'\rangle$ ist der Realteil von $\langle x, x'\rangle$; es gilt $|\mathsf{R}\langle x, x'\rangle| \leq |\langle x, x'\rangle| \leq |\mathsf{R}\langle x, x'\rangle| + |\mathsf{R}\langle ix, x'\rangle|$). Die reellen Halbnormen $(*)$ sind offenbar monoton auf K (für die durch K induzierte Ordnungsstruktur), womit nach (1.1) die Behauptung bewiesen ist. Für $E' = H' - H'$ brauchen wir nur zu bemerken, daß die Halbnormen $(*)$ (jetzt für $x' \in H'$) auf $K - iK$ monoton sind. Hieraus folgt, daß $K - iK$, also auch $K + iK$ normal ist für $\sigma(E, E')$.

Aus dem Charakter des vorstehenden Beweises entnimmt man, daß auch die Dualitätsrelation (1.5) im komplexen Fall bestehen bleibt. Sei $\langle F, G\rangle$ ein komplexes Dualsystem, $\mathfrak{S}$ eine gesättigte Klasse (schwach) beschränkter Teilmengen von G. Da $x \to |\mathsf{R}\langle x, x'\rangle|$, wie eben bemerkt wurde, eine zu $x \to |\langle x, x'\rangle|$ auf F äquivalente Halbnorm ist (und zwar gleichmäßig in x'), kann die $\mathfrak{S}$-Topologie auf F durch das System

$$x \to p_S(x) = \sup_{x' \in S} |\mathsf{R}\langle x, x'\rangle| \qquad (S \in \mathfrak{S})$$

reeller Halbnormen definiert werden. Ist daher K' ein $\mathfrak{S}$-Kegel in G, so kann in $p_S(x)$ S durch $S \cap K'$ ersetzt werden. Hieraus folgt unmittelbar, daß K normal ist für die $\mathfrak{S}$-Topologie auf F. Der zweite Teil des Beweises von (1.5) überträgt sich ebenfalls mit geringer Modifikation. Damit erhalten wir

(6.4). *Es sei $\langle F, G\rangle$ ein komplexes Dualsystem.*

a) *Ist K (bzw. $K + iK$) normal für eine zulässige[5]) $\mathfrak{S}$-Topologie auf F, so ist K' (bzw. H') ein strikter $\mathfrak{S}$-Kegel in G.*

b) *Ist K' (bzw. H') $\mathfrak{S}$-Kegel in G, so ist K (bzw. $K + iK$) normal für die $\mathfrak{S}$-Topologie auf F.*

Ist E (bzw. $E[\mathfrak{T}]$) ein reeller linearer (bzw. reeller topologischer linearer) Raum, so ist E nicht immer assoziiert zu einem komplexen (bzw. komplexen topologischen) linearen Raum. Um E in einen komplexen Raum einzubetten (was z. B. für Zwecke der Spektraltheorie häufig erforderlich ist), bildet man gewöhnlich das Produkt $E \times E$: $(x, y) \to (-y, x)$ ist ein (für einen topologischen linearen Raum E topologischer) Automorphismus φ von $E \times E$ (in der Produkt-

[5]) d. h. mit der Dualität zwischen F und G verträgliche.

topologie), mit $\varphi^2 = -1$. Wegen $i(x, 0) = (0, x)$ (s. o.) schreibt sich (x, y) eindeutig als $x + iy$, d. h. $(E \times E)_0 = E \oplus iE$ ist (topologische) direkte Summe der reellen Teilräume E und iE von $E \times E$. Wir bezeichnen den komplexen Raum F, dessen assoziierter reeller Raum $E \oplus iE$ ist, als *komplexe Erweiterung* von E. Beim Übergang von E zu seiner komplexen Erweiterung ergibt sich folgender Sachverhalt:

(6.5). *Sei F komplexe Erweiterung des reellen lokalkonvexen Raumes E. Ist E durch einen Kegel K halbgeordnet (bzw. regulär halbgeordnet[6])), so ist F sowohl für die durch K, als auch für die durch $K + iK$ erzeugte Ordnungsstruktur ein halbgeordneter (bzw. regulär halbgeordneter) lokalkonvexer Raum. Ist K ein normaler (bzw. ein BZ-) Kegel in E, so ist $K + iK$ normal (bzw. BZ-Kegel) in F. Jede stetige Linearform auf F ist bei normalem K bezüglich der K-Halbordnung (bzw. der $K + iK$-Halbordnung) in stark (bzw. schwach) positive stetige Komponenten zerlegbar.*

Beweis. Wenn K in $E[\mathfrak{T}]$ normal ist, so ist $K + iK = K \times K$ nach (1.2) in $E[\mathfrak{T}] \times E[\mathfrak{T}] = F$ normal. Hieraus folgt die zweite, und mit Hilfe von Def. 3 und (1.7), 4° auch die erste Behauptung. Ist $\{B_\alpha\}$ ein Fundamentalsystem beschränkter Mengen in E, so ist $\{B_\alpha \times B_\alpha\}$ ein solches System in $E[\mathfrak{T}] \times E[\mathfrak{T}]$, also auch in F, und mit K in E ist $K + iK$ BZ-Kegel in F. Die übrigen Aussagen folgen nun aus (6.1) und (6.2).

Es seien E, F komplexe lineare Räume, und es sei F komplexe Erweiterung eines reellen Raumes F_1, d. h. es sei $F_0 = F_1 \oplus iF_1$. E sei durch einen Kegel K, F_1 durch K_1 und F durch $K_1 + iK_1$ halbgeordnet. Ist $x \to y_1 + iy_2$ eine lineare Abbildung von E in F, so kann man (wie bei den Linearformen) fordern, daß $x \in K$ entweder $y_1 \in K_1$ oder $y_1 \in K_1$ *und* $y_2 \in K_1$ nach sich zieht, und entsprechend zwischen schwach und stark positiven linearen Abbildungen unterscheiden. Ist E selbst eine komplexe Erweiterung, etwa $E_0 = E_1 \oplus iE_1$, K_2 ein Kegel in E_1 und $K = K_2 + iK_2$, so leuchtet unmittelbar ein, daß die Fortsetzung einer positiven linearen Abbildung von E_1 in F_1 auf E (und in F) stark positiv ist.

7. Konvergenzsätze

Wir stellen in diesem Abschnitt eine Reihe von Sätzen auf, die Konvergenzeigenschaften gerichteter Systeme für eine gegebene lokalkonvexe Topologie $\mathfrak{T}$ in einem halbgeordneten linearen Raum E aussagen. Die betrachteten Räume $E[\mathfrak{T}]$ können reell oder komplex sein.

Eine für die Ordnungsrelation „$\leq$" (bzw. „$\geq$") gerichtete Menge[7]) ist eine Teilmenge H von E, die zu jedem Paar $x \in H$, $y \in H$ ein z mit $z \geq x$, $z \geq y$ (bzw. $z \leq x$, $z \leq y$) enthält. Jede einem Element $x_0 \in H$ zugeordnete Teilmenge $H' = \{x \in H: x_0 \leq x\}$ (bzw. $\{x \in H: x \leq x_0\}$) heißt ein *Ende* von H. Ist H eine nichtleere gerichtete Menge, so bildet die Menge $\{H'\}$ aller Enden von H eine Filterbasis in E; den zugehörigen Filter bezeichnen wir kurz als den Filter der

---

[6]) Def. 3, HV p. 126.

[7]) Ensemble filtrant; „gerichtetes System" wird vielfach für Moore-Smith-Folgen gebraucht, für die wir jedoch keine Verwendung haben.

Enden von H. Eine Folge in E, $\{x_n\}$, heiße *quasimonoton*, wenn es zu jedem $k \in N$ ein $n = n(k)$ gibt, so daß $x_m \geq x_k$ (bzw. $x_m \leq x_k$) gilt für alle $m \geq n(k)$. Eine quasi-monotone Folge ist ein spezielles Beispiel einer gerichteten Menge.— Ein Filter in einem topologischen linearen Raum heißt *beschränkt*, wenn er eine beschränkte Menge enthält. Schließlich nennen wir einen Filter in einem halbgeordneten linearen Raum *positiv*, wenn eines seiner Elemente eine Teilmenge des positiven Kegels K ist.

Wir notieren zunächst folgenden bekannten Satz.

(7.1). *Es sei $L[\mathfrak{T}]$ ein topologischer linearer Raum, halbgeordnet durch einen abgeschlossenen Kegel K. Ist H eine gerichtete Menge, für die der Filter der Enden von H gegen ein $x \in L$ konvergiert, so gilt $x = \sup H$.*

Für einen Beweis s. [9], p. 26 prop. 6. Wir kommen jetzt zu einem Satz, der als Verallgemeinerung eines bekannten Satzes von Dini über die Konvergenz monotoner Funktionenfolgen aufgefaßt werden kann.

(7.2). *Sei $E[\mathfrak{T}]$ lokalkonvex und halbgeordnet durch einen normalen Kegel K. Sei H eine gerichtete Teilmenge von E, für die der Filter $\mathfrak{F}(H)$ der Enden von H schwach konvergiert. Dann konvergiert $\mathfrak{F}(H)$ für $\mathfrak{T}$.*

Beweis. Wir können ohne Einschränkung der Allgemeinheit voraussetzen, daß H nichtleer, ferner absteigend (d. h. für „$\geq$") gerichtet ist und $\mathfrak{F}(H)$ schwach gegen 0 konvergiert.

Wir nehmen an, daß $\mathfrak{F}(H)$ für $\mathfrak{T}$ nicht gegen 0 konvergiert. Dann gibt es einen Filter $\mathfrak{F}_1$, der feiner ist als $\mathfrak{F}$, und eine $\mathfrak{T}$-Nullumgebung U in E, so daß $F_1 \cap U$ leer ist für ein $F_1 \in \mathfrak{F}_1$. Wir denken uns U als Element einer der Def. 1 (in E_0) genügenden Nullumgebungsbasis. $H_x = \{z \in H : z \leq x\}$ bezeichne das von $x \in H$ erzeugte Ende von H. Da $\mathfrak{F}_1$ feiner als $\mathfrak{F}$ ist, hat jedes Ende H_x einen nichtleeren Durchschnitt mit F_1. Nach dem Auswahlaxiom werde jeder der Mengen $H_x \cap F_1 (x \in H)$ eines ihrer Elemente, etwa y_x, zugeordnet. Die Menge $\{y_x\}_{x \in H}$ ist selbst gerichtet für „$\geq$", denn aus $z \leq y_u$, $z \leq y_v$ folgt $y_z \leq y_u$, $y_z \leq y_v$. (Da H gerichtet ist, existiert für je zwei Elemente y_u, y_v ein solches z.)

Die Menge

$$C = \bigcup_{x \in H} (y_x + K)$$

ist konvex. Denn ist $y_z \leq y_x$, so ist $y_x + K = y_z + (y_x - y_z) + K \subset y_z + K$. Da $\{y_x\}_{x \in H}$ gerichtet ist, ist C konvex. Nun ist mit K auch die $\mathfrak{T}$-abgeschlossene Hülle $\bar{K}$ ein normaler, folglich echter Kegel [vgl. (1.1), (1.2)]. Nach (7.1) ist, da $\mathfrak{F}(H)$ schwach gegen 0 konvergiert, $0 = \inf H$ für die durch $\bar{K}$ erzeugte Ordnungsstruktur. Es ist also $H \subset \bar{K}$; aus Def. 1 folgt nun nach Wahl von U und wegen $F_1 \cap U = \emptyset$, $y_x \in F_1$, daß $(y_x + K) \cap U$ leer ist für jedes $x \in H$; folglich ist auch $C \cap U = \emptyset$. Es gibt also eine reelle abgeschlossene Hyperebene in $E[\mathfrak{T}]$, die C und U trennt. Dies steht im Widerspruch zur schwachen Konvergenz von $\mathfrak{F}$ gegen 0 (denn jedes Ende von H enthält ein Element von C), und der Satz ist bewiesen.

Corollar. *In einem halbgeordneten lokalkonvexen Raum, dessen positiver Kegel normal ist, ist jede quasi-monotone und schwach konvergente Folge konvergent.*

Beispiel. In jedem der mit der natürlichen Halbordnung versehenen Banachräume a) bis e) (HV, p. 130) ist der positive Kegel normal. Daher ist in jedem dieser Räume insbesondere jede monotone, schwach konvergente Folge normkonvergent. Hierin ist der klassische Konvergenzsatz von Dini enthalten:

Jede monotone Folge $\{f_n\}$ *auf* $[0, 1]$ *stetiger reeller Funktionen, die punktweise gegen ein stetiges* f *konvergiert, ist gleichmäßig konvergent auf* $[0, 1]$.

Denn aus der Voraussetzung folgt, daß $f_n \to f$ gilt für die schwache Topologie auf $[0, 1]$ (vgl. Banach [1], p. 134). Eine allgemeinere Fassung (vgl. [5′], p. 52 th. 1) lautet:

Es sei H *eine gerichtete Menge, für die Beziehung „$\leqq$" (bzw. „$\geqq$"), von stetigen reellen Funktionen auf einem lokalkompakten Raum* X. *Ist die obere (bzw. untere) Einhüllende* g *von* H *(endlich und) stetig auf* X, *so kann* g *auf jeder kompakten Teilmenge von* X *gleichmäßig durch Funktionen aus* H *approximiert werden*[8]).

Wir beweisen als nächstes einen Satz, der in gewissem Sinne eine Umkehrung von (7.1) ist.

(7.3). *Sei* $E[\mathfrak{T}]$ *ein halbreflexiver lokalkonvexer Raum, halbgeordnet durch einen normalen Kegel* K. H *sei eine für „$\leqq$" (bzw. „$\geqq$") gerichtete Teilmenge von* E, *die einer der Bedingungen genügt:*

a) H *besitzt eine obere (bzw. untere) Schranke in* E.

b) H *ist beschränkt (als Teilmenge von* $E[\mathfrak{T}]$*).*

Dann konvergiert der Filter der Enden von H *gegen ein* $x \in E$, *und es gilt* $x = \sup H$ *(bzw.* $x = \inf H$*) für die durch* $\bar{K}$ *erzeugte Ordnungsstruktur.*

Beweis. Wir bemerken, daß die Aussage des Satzes von dem benutzten Skalarkörper ($\mathbf{R}$ oder $\mathbf{C}$) nicht abhängt. Wir setzen daher E als reellen Raum voraus und führen den Beweis zunächst für „$\leqq$" und unter der Voraussetzung b).

Wir definieren eine endliche reelle Funktion $x' \to f(x')$ auf K', indem wir für $x' \in K'$

$$f(x') = \sup_{x \in H} \langle x, x' \rangle$$

setzen. Offenbar ist f positiv homogen. Ferner ist f additiv; denn einerseits ist $f(x' + y') \leqq f(x') + f(y')$; andererseits sei $\langle x_1, x' \rangle > f(x') - \frac{1}{2}\varepsilon$ und $\langle x_2, y' \rangle > f(y') - \frac{1}{2}\varepsilon$ für gegebenes $\varepsilon > 0$ und geeignete $x_1 \in H$, $x_2 \in H$. Da H gerichtet ist, existiert $z \in H$ mit $z \geqq x_1$, $z \geqq x_2$, woraus $\langle z, x' + y' \rangle = \langle z, x' \rangle + \langle z, y' \rangle > f(x') + f(y') - \varepsilon$ folgt; damit ist bewiesen, daß f auf K' additiv ist. Weil K normal ist, gilt nach (1.3) $E' = K' - K'$ und f setzt sich in eindeutiger Weise zu einer Linearform auf E' fort, die wir ebenfalls mit f bezeichnen.

Aus der Definition von f folgt, daß der Filter der Enden von H einfach (d. h. punktweise) auf K' gegen die Abbildung $x' \to f(x')$ konvergiert; nach dem beim Beweis der Additivität verwendeten Muster zeigt man — H ist gerichtet und $E' = K' - K'$ —, daß dieser Filter linearer Abbildungen von E' in $\mathbf{R}$ auf ganz E' einfach gegen f konvergiert. Nun ist E halbreflexiv, also E'

[8]) Ist X im Unendlichen abzählbar, so existiert eine Folge $\{f_n\}$, $f_n \in H$, die, gleichmäßig auf jeder kompakten Teilmenge von X, gegen f konvergiert.

tonneliert für die starke Topologie $\beta(E', E)$. Daher ergibt sich aus dem Banach-Steinhaus-Theorem ([8], p. 27 th. 2 cor.), daß der Filter $\mathfrak{F}(H)$ der Enden von H punktweise in E' gegen eine für $\beta(E', E)$ stetige Linearform konvergiert. Jede solche Linearform schreibt sich aber $x' \to \langle x, x'\rangle$ für ein $x \in E$, wir haben also $f(x') = \langle x, x'\rangle$. Da die einfache Konvergenz von $\mathfrak{F}(H)$ gegen f nichts anderes als die Konvergenz gegen x für $\sigma(E, E')$ ist, ergibt sich mit Hilfe von (7.2) (K ist normal) nun $\lim \mathfrak{F}(H) = x$ für $\mathfrak{T}$. $x = \sup H$ für die durch $\bar{K}$ erzeugte Halbordnung ist eine unmittelbare Folge von (7.1).

Entsprechend verläuft der Beweis, wenn H für „$\geqq$“ gerichtet ist ($-H$ ist dann gerichtet für „$\leqq$“). Wir haben noch zu zeigen, daß die Voraussetzung a) ebenfalls zum Beweis hinreicht. Es sei z eine obere Schranke für H, d. h. $x \leqq z$ für alle $x \in H$. Sei $x_0 \in H$ und H_0 das von x_0 erzeugte Ende von H. Für jedes $x \in H_0$ gilt $x_0 \leqq x \leqq z$; folglich ist die Menge H_0 wegen der Normalität von K nach (1.1), zweites Corollar, beschränkt in $E[\mathfrak{T}]$. Daher erfüllt H_0 die Voraussetzung b); nach dem vorstehenden Beweis konvergiert der Filter $\mathfrak{F}(H_0)$ der Enden von H_0 für $\mathfrak{T}$ gegen ein $x \in E$. Folglich konvergiert der feinere Filter $\mathfrak{F}(H)$ erst recht gegen x, und es ist wieder $x = \sup H$ für die durch $\bar{K}$ erzeugte Ordnungsstruktur. Damit ist (7.3) bewiesen.

Wir formulieren noch eine vereinfachte Form von (7.3).

Corollar. *E sei ein reflexiver Banachraum, K ein normaler abgeschlossener Kegel in E. Jede (für die durch K induzierte Ordnungsstruktur) quasi-monotone Folge $\{x_n\}$ mit $\sup_n \|x_n\| < \infty$ konvergiert; es gilt $\lim_n x_n = \sup_n x_n$.*

Zu diesem Corollar vergleiche man einen Satz von Amemiya [2′], s. auch Yamamuro [16′].

Beispiel. Der positive Kegel K' im (reflexiven) Raum $\mathscr{D}'$ der Schwartzschen Distributionen (d. h. der Kegel der positiven Masse auf $\mathbf{R}^n$, HV p. 131, Beispiel 2) ist normal und abgeschlossen in $\mathscr{D}'$. Daher folgt aus (7.3):

Es sei H eine aufsteigend gerichtete, majorisierte (bzw. beschränkte) Menge von Distributionen. Der Filter der Enden von H konvergiert gegen eine Distribution f, und es gilt $f = \sup H$.

Wir beweisen noch drei weitere Aussagen, von denen (7.4) und (7.6) gleichzeitig Nachträge zu Abschnitt 1 sind.

(7.4). *Es sei K ein echter Kegel in dem lokalkonvexen Raum E[*]. Ist x_0 ein innerer Punkt von K, so ist der Durchschnitt der reellen Hyperebene $\{x' : \mathsf{R}\langle x_0, x'\rangle = 1\}$ mit K' kompakt für $\sigma(E', E)$.*

Beweis. Die Menge der Linearformen in diesem Durchschnitt ist gleichstetig, da diese x' auf $\{x : 0 \leqq x \leqq x_0\} = M$ („$\leqq$“ die durch K erzeugte Halbordnung) gleichmäßig beschränkt sind. (Denn ist x_0 innerer Punkt von K, so ist $\frac{1}{2} x_0$ innerer Punkt von M.) Daraus folgt die Behauptung unmittelbar.

(7.5). *E sei ein lokalkonvexer Raum, halbgeordnet durch einen Kegel K mit inneren Punkten. Jeder (für die durch K' erzeugte Ordnungsstruktur) positive Filter in E', der für $\sigma(E', E)$ gegen 0 konvergiert, konvergiert auch für $\beta(E', E)$.*

---

[*]) „E lokalkonvex“ ist nicht wesentlich.

Beweis. Sei $\mathfrak{F}$ ein positiver Filter in E', und $\lim \mathfrak{F} = 0$ für $\sigma(E', E)$. Es sei U eine beliebige Nullumgebung in E' für $\beta(E', E)$. Ist x_0 ein innerer Punkt von K, so ist nach (7.4) die Menge $S_\varepsilon = K' \cap \{x' : \mathrm{R}\langle x_0, x'\rangle \leq \varepsilon\}$ kompakt für $\sigma(E', E)$. Da jede schwach kompakte Teilmenge von E' stark beschränkt ist, gilt $S_\varepsilon \subset U$ für geeignetes $\varepsilon > 0$. Wegen der schwachen Konvergenz von $\mathfrak{F}$ gegen 0 existiert $F_1 \in \mathfrak{F}$ mit $\mathrm{R}\langle x_0, x'\rangle \leq \varepsilon$ für $x' \in F_1$. Andererseits gibt es ein $F_2 \in \mathfrak{F}$ mit $F_2 \subset K'$ nach Voraussetzung. Daher gilt $F_1 \cap F_2 \subset U$, woraus die Behauptung folgt.

Daß ein normaler Kegel in einem lokalkonvexen Raum nur unter sehr einschränkenden Bedingungen innere Punkte haben kann, zeigt

(7.6). *Existiert in $E[\mathfrak{T}]$ ein schwach (d. h. für $\sigma(E, E')$) normaler Kegel mit nichtleerem Inneren, so kann $\mathfrak{T}$ durch eine einzige Norm definiert werden.*

Beweis. Sei $\mathring{K} \neq \emptyset$ das Innere von K, und $x_0 \in \mathring{K}$. Die Menge $\mathring{K} \cap (x_0 - \mathring{K})$ ist konvex, offen und nicht leer (da sie $\frac{1}{2}x_0$ enthält). Andererseits ist diese Menge nach (1.1), zweites Corollar, beschränkt (denn die beschränkten Mengen sind für $\mathfrak{T}$ und $\sigma(E, E')$ dieselben), woraus die Behauptung nach einem bekannten Satz von Kolmogoroff folgt.

II. Teil: Positive lineare Abbildungen

8. Halbgeordnete Räume linearer Abbildungen

Es seien E, F lokalkonvexe lineare Räume über dem gleichen Skalarkörper ($\mathbf{R}$ oder $\mathbf{C}$). Die Gesamtheit stetiger linearer Abbildungen von E in F bildet einen linearen Raum $\mathfrak{L}(E, F)$. Es sei $\mathfrak{S}$ eine gesättigte Klasse beschränkter Teilmengen von E, $\{p_\alpha\}_{\alpha \in A}$ eine Familie von Halbnormen, die die Topologie von F erzeugt. Die durch die Halbnormen

$$T \to p_{\alpha, S}(T) = \sup_{x \in S} p_\alpha(Tx) \qquad (S \in \mathfrak{S},\ \alpha \in A)$$

auf $\mathfrak{L}$ erzeugte lokalkonvexe Topologie heißt Topologie der gleichmäßigen Konvergenz auf (den Mengen von) $\mathfrak{S}$ oder kurz $\mathfrak{S}$-Topologie[10]. Der mit der $\mathfrak{S}$-Topologie versehene Raum $\mathfrak{L}(E, F)$ werde mit $\mathfrak{L}_{\mathfrak{S}}(E, F)$ oder einfach $\mathfrak{L}_{\mathfrak{S}}$ bezeichnet, wenn keine Verwechslungen möglich sind.

Sind E, F lokalkonvexe Räume und K bzw. H Kegel in E bzw. F, so bilden die Abbildungen $T \in \mathfrak{L}$ mit $T(K) \subset H$ einen Kegel $\mathfrak{K}$ in $\mathfrak{L}$.

(8.1). *Ist K total in E und H (bzw. $\overline{H}$) echt in F, so ist $\mathfrak{L}(E, F)$ durch $\mathfrak{K}$ halbgeordnet (bzw. regulär halbgeordnet). Ist H abgeschlossen in F, so ist $\mathfrak{K}$ abgeschlossen in $\mathfrak{L}_{\mathfrak{S}}$ für jede $\mathfrak{S}$-Topologie.*

Beweis. Wir beweisen zunächst die zweite Behauptung. Nach Definition der $\mathfrak{S}$-Topologie ist die Abbildung $(x, T) \to Tx$ von $E \times \mathfrak{L}_{\mathfrak{S}}$ in F getrennt stetig. Daher ist für jedes feste $x \in E$ die Abbildung $f_x : T \to Tx$ von $\mathfrak{L}_{\mathfrak{S}}$ in F stetig. Folglich ist $f_x^{-1}(\overline{H})$ abgeschlossen in $\mathfrak{L}_{\mathfrak{S}}$, wenn H in F abgeschlossen ist. Da $\mathfrak{K}$ der Durchschnitt dieser Mengen für alle $x \in K$ ist, folgt aus der Abgeschlossenheit von H diejenige von $\mathfrak{K}$.

---

[10]) Vgl. [8], chap. III und [14]. Wir nehmen stets an, daß die $\mathfrak{S}$-Topologie Hausdorffsch ist.

Es sei jetzt K total und H echt, und es sei $T \in \mathfrak{K} \cap (-\mathfrak{K})$. Für $x \in K$ folgt $y = Tx \in H$ und $-y \in H$, also $y = 0$. Da K total ist, ist $T = 0$, daher $\mathfrak{K}$ echt. $\mathfrak{L}$ ist regulär halbgeordnet nach Def. 3, wenn $\overline{\mathfrak{K}}$ echt ist in $\mathfrak{L}_{\mathfrak{S}}$. Nun folgt aus der Stetigkeit von f_x (s. o.), daß $Tx \in \overline{H}$ gilt für jedes $T \in \overline{\mathfrak{K}}$ und $x \in K$. Ist also $\overline{H}$ echt, so folgt aus $T \in \overline{\mathfrak{K}} \cap (-\overline{\mathfrak{K}})$ wieder $T = 0$, w. z. b. w.

(8.2). *Es seien E, F reell und es sei K nicht dicht in E. Ist $\mathfrak{K}$ in $\mathfrak{L}$ (bzw. $\overline{\mathfrak{K}}$ in $\mathfrak{L}_{\mathfrak{S}}$) echt, so ist F durch H halbgeordnet (bzw. regulär halbgeordnet).*

Beweis. Wir haben zu zeigen, daß H (bzw. $\overline{H}$) ein echter Kegel in F ist. Da K nicht dicht ist, ist $\overline{K} \neq E$. Andererseits ist $\overline{K}$ konvex, folglich gibt es eine (reelle) abgeschlossene homogene Hyperebene L in E, so daß $\overline{K}$ auf einer Seite von L liegt. (Denn ist U eine offene konvexe Menge mit $U \cap \overline{K} = \emptyset$, so ist auch $\overline{K} \cap \left\{\bigcup_{\lambda>0} \lambda U\right\} = \emptyset$, da K 0 enthält.)

Wir zeigen, daß L nicht $\overline{K}$ enthält. Wäre dies der Fall, so wäre jedes $T \in \mathfrak{L}(E, F)$, das auf L verschwindet, in $\mathfrak{K}$ enthalten. Es sei $z \notin L$; dann ist $E = [z] \oplus L$, wo $[z]$ den von z aufgespannten linearen Teilraum von E bezeichnet. Jede lineare Abbildung von $[z]$ in F läßt sich stetig auf E fortsetzen, indem den Elementen von L die 0 in F zugeordnet wird. Daher ist jede solche Fortsetzung in $\mathfrak{K}$, was der vorausgesetzten Echtheit von $\mathfrak{K}$ widerspricht.

Es gibt also ein $x_0 \in \overline{K}$, für das $f(x_0) = 1$ gilt, wenn $f(x) = 0$ eine Gleichung von L ist. Wegen der Stetigkeit von f gibt es ein $y_0 \in K$ mit $f(y_0) = 1$. Es gilt $E = [y_0] \oplus L$, und die Projektion $x \to f(x) y_0$ ist stetig. Wir nehmen nun an, H wäre nicht echt. Dann gilt für ein $0 \neq w \in F$: $w \in H \cap (-H)$. Daraus folgt, daß die stetigen Abbildungen $x \to f(x) w$ und $x \to -f(x) w$ beide in $\mathfrak{K}$ sind, was der Echtheit von $\mathfrak{K}$ widerspricht.

Wäre $\overline{\mathfrak{K}}$ in $\mathfrak{L}_{\mathfrak{S}}$ echt, aber $\overline{H}$ nicht echt in F, so erhielten wir zwei Abbildungen $x \to f(x) w_0$ und $x \to -f(x) w_0$, mit $w_0 \in \overline{H} \cap (-\overline{H})$ und $f \in K'$. Wir betrachten die Spur eines gegen w_0 konvergenten Filters auf H. Vermöge der Zuordnung $w \to T_w$, $T_w(x) = f(x) w$, entspricht diesem Filter ein Filter in $\mathfrak{L}$, der für jede $\mathfrak{S}$-Topologie gegen die Abbildung $x \to f(x) w_0$ konvergiert. Es ergäbe sich also, daß $\overline{\mathfrak{K}}$ in $\mathfrak{L}_{\mathfrak{S}}$ nicht echt ist, entgegen der Voraussetzung, w. z. b. w.

Bemerkung. Sind E, F sowie K, H komplexe Erweiterungen reeller Räume (bzw. Kegel)[11]), so bleibt (8.2) bestehen.

Wir kommen jetzt zu einer Bedingung für die Normalität des Kegels $\mathfrak{K}$ in $\mathfrak{L}_{\mathfrak{S}}(E, F)$.

(8.3). *Es sei K ein $\mathfrak{S}$-Kegel in E und H normal in F. Dann ist $\mathfrak{K}$ normal für die $\mathfrak{S}$-Topologie auf $\mathfrak{L}(E, F)$.*

Beweis. Nach (1.1) genügt es zu zeigen, daß die $\mathfrak{S}$-Topologie auf $\mathfrak{L}$ durch eine Familie von Halbnormen definiert wird, die auf $\mathfrak{K}$ (für die durch $\mathfrak{K}$ in $\mathfrak{L}$ erzeugte Ordnungsstruktur) monoton sind. Nach (8.1) bemerken wir zunächst, daß $\mathfrak{L}$ durch $\mathfrak{K}$ regulär halbgeordnet ist (denn für einen normalen Kegel $H \subset F$ ist $\overline{H}$ normal und folglich echt). Da K $\mathfrak{S}$-Kegel ist, bilden nach Def. 2 und 6 die abgeschlossenen, konvexen und symmetrischen Hüllen der Durchschnitte

[11]) Vgl. Abschnitt 6, p. 263/264.

$S \cap K$, $S \in \mathfrak{S}$, ein Fundamentalsystem für $\mathfrak{S}$ [vgl. (1.4)]. Folglich definieren die Halbnormen

$$T \to q_{\alpha,S}(T) = \sup_{x \in S \cap K} p_\alpha(Tx) \qquad (\alpha \in A,\ S \in \mathfrak{S})$$

die $\mathfrak{S}$-Topologie in $\mathfrak{L}$, wenn $\{p_\alpha\}$ die Topologie von F erzeugen[12]). Da H normal ist, können wir nach Def. 6 und (1.1) die p_α als auf H monoton voraussetzen, woraus die Monotonie der $q_{\alpha,S}$ auf $\mathfrak{K}$ unmittelbar folgt.

(8.3) ist offenbar eine Verallgemeinerung des ersten Teiles von (1.5). Die Verallgemeinerung des Zerlegungssatzes in (1.5) auf allgemeinere Räume linearer Abbildungen stößt dagegen auf Schwierigkeiten. Wir beweisen zunächst eine notwendige Bedingung für die Zerlegbarkeit von $\mathfrak{L}$ nach $\mathfrak{K}$, d. h. für die Gültigkeit der Beziehung $\mathfrak{L} = \mathfrak{K} - \mathfrak{K}$.

(8.4). *Es sei $\mathfrak{L} = \mathfrak{K} - \mathfrak{K}$. Dann ist $F = \overline{H} - \overline{H}$ und K normal in E für die schwache Topologie $\sigma(E, E')$.*

Beweis. Wir versehen $\mathfrak{L}(E, F)$ mit der Topologie der einfachen Konvergenz. Dann existiert, wie man weiß ([8], p. 77), ein algebraischer Isomorphismus des Tensorproduktes $E \otimes F'$ auf den Dual $\mathfrak{L}'$ von $\mathfrak{L}$, vermöge dessen dem Tensorprodukt $x \otimes y'$ die (für die Topologie der einfachen Konvergenz auf $\mathfrak{L}$ stetige) Linearform $T \to \langle Tx, y'\rangle$ entspricht. Es sei nun $\mathsf{R}\langle Tx, y'\rangle \geqq 0$ für alle $x \in K$, $y' \in H'$ und ein gewisses $T \in \mathfrak{L}$. Hieraus folgt $T \in \overline{\mathfrak{K}}$; denn wäre $Tx_0 \notin \overline{H}$ für ein $x_0 \in K$, so könnten die Mengen $\overline{H}$ und $\bigcup_{\lambda > 0} \lambda U$ (U eine offene, konvexe Umgebung von Tx_0, deren Durchschnitt mit $\overline{H}$ leer ist) durch eine reelle abgeschlossene Hyperebene $\{y : \mathsf{R}\langle y, y_0'\rangle = 0\}$, $y_0' \in H'$, in F getrennt werden, und es wäre $\mathsf{R}\langle Tx_0, y_0'\rangle < 0$ entgegen der Voraussetzung.

Daher besteht, wenn wir $\mathfrak{L}'$ mit $E \otimes F'$ identifizieren, der Kegel $\mathfrak{K}'$ in $\mathfrak{L}'$ aus der schwach abgeschlossenen Hülle der Menge aller Elemente der Gestalt $\sum_{i=1}^{n} x_i \otimes y_i'$ mit $x_i \in K$, $y_i' \in H'$, $n \in \mathsf{N}$. Weiter ist für jedes feste $x_0 \neq 0$ in E der lineare Teilraum von $\mathfrak{L}'$

$$\{x_0 \otimes y' : y' \in F'\}$$

zu F' in der schwachen Topologie $\sigma(F', F)$ isomorph, wenn $\mathfrak{L}'$ mit $\sigma(\mathfrak{L}', \mathfrak{L})$ versehen wird; und eine entsprechende Aussage gilt für $\{x \otimes y_0' : x \in E\}$ und E, wenn y_0' fest und $\neq 0$ ist. Denn $x \otimes y' = 0$ bedeutet $\langle Tx, y'\rangle = 0$ für alle $T \in \mathfrak{L}$, so daß mindestens eine der Beziehungen $x = 0$ oder $y' = 0$ gilt. Die Homöomorphie bezüglich der schwachen Topologien ist leicht nachzuprüfen.

Aus $\mathfrak{L} = \mathfrak{K} - \mathfrak{K}$ und (1.6) bzw. (6.3) folgt, daß $\mathfrak{K}' \subset \mathfrak{L}'$ für $\sigma(\mathfrak{L}', \mathfrak{L})$ normal ist. Nach dem eben Gesagten ist aber sowohl K wie H' isomorph zu einem Teilkegel von $\mathfrak{K}'$, woraus die Normalität von K für $\sigma(E, E')$ bzw. von H' für $\sigma(F', F)$ hervorgeht. Nach (1.6) bzw. (6.3) ist also $F = \overline{H} - \overline{H}$. Damit ist (8.4) bewiesen.

---

[12]) Falls E, F (und folglich $\mathfrak{L}$) komplex sind, brauchen die p_α nicht notwendig komplexe Halbnormen zu sein.

Ist umgekehrt K schwach normal in E und $F = \overline{H} - \overline{H}$, so ist $\mathfrak{R} - \mathfrak{R}$ dicht in $\mathfrak{L}(E, F)$ für die Topologie der einfachen Konvergenz[13]).

Denn ist $\langle x \otimes y', T \rangle = 0$ für alle T, die vermöge der kanonischen Einbettung von $E' \otimes F$ in $\mathfrak{L}(E, F)$ den Tensorprodukten $x' \otimes y$, $x' \in K'$, $y \in \overline{H}$ entsprechen (das sind die Abbildungen $x \to \langle x, x' \rangle y$ vom Rang 1), so folgt wegen $E' = K' - K'$ $x = 0$ oder $y' = 0$, also die Behauptung.

Es bleibt die Frage nach der Zerlegbarkeit einer stetigen Abbildung T in positive Komponenten, d. h. nach einer Darstellung $T = T_1 - T_2$ mit $T_i \in \mathfrak{R}$ $(i = 1, 2)$. Es ist klar, daß ein gegebenes $T \in \mathfrak{L}$ eine solche Zerlegung genau dann besitzt, wenn es zu der auf $E \times F'$ getrennt schwachstetigen Bilinearform $(x, y') \to f(x, y') = \langle T x, y' \rangle$ eine ebensolche Bilinearform f_1 gibt, für die aber auf $K \times H'$ die Beziehung $f_1(x, y') \geq \sup\{0, f(x, y')\}$ gilt [die Topologie von E sei $\tau(E, E')$]. Wegen der (algebraischen) Isomorphie von $\mathfrak{L}'_{\mathfrak{S}}$ (für die $\mathfrak{S}$-Topologie der einfachen Konvergenz) zu $E \otimes F'$ läuft das Problem $\mathfrak{L} = \mathfrak{R} - \mathfrak{R}$ nach (1.3), (1.6) darauf hinaus, die Normalität des konvexen Kegels conv $\overline{K} \otimes H'$ für eine Topologie auf $E \otimes F'$ nachzuweisen, für die der Dual von $E \otimes F'$ mit $\mathfrak{L}$ übereinstimmt.

Die allgemeinste positive Antwort auf das Zerlegungsproblem ergibt sich, sobald der Kegel conv $\overline{K} \otimes H'$ für die induktive Topologie des Tensorproduktes ([7'] I, p. 73f.) als normal nachgewiesen werden kann[14]). Normalität dieses Kegels für die projektive Topologie (l. c., p. 32) zieht die Zerlegbarkeit jeder Abbildung T nach sich, die eine Nullumgebung in E in eine schwachkompakte Teilmenge von F abbildet.

(8.5). *Es sei E ein beliebiger, F ein quasi-vollständiger lokalkonvexer Raum*[13]). *K sei ein normaler Kegel in E, H ein strikter BZ-Kegel in F. Jede nukleare Abbildung von E in F ist Differenz zweier nuklearer Abbildungen, die K in $\overline{H}$ abbilden.*

Beweis. Jede nukleare Abbildung von E in F schreibt sich $\sum_{n=1}^{\infty} \lambda_n x'_n \otimes y_n$, ([7'] I, p. 83), wo $\{\lambda_n\}$ eine summierbare Folge, $\{x'_n\}$ eine gleichstetige Folge in E', und $\{y_n\}$ eine Folge aus einer absolutkonvexen beschränkten Teilmenge von F ist. Da K als normal vorausgesetzt ist, gilt $x'_n = u'_n - v'_n$ $(n \in \mathrm{N})$, wo die u'_n, v'_n Elemente von K' sind und als gleichstetig angenommen werden können. [Dies ergibt sich durch Anwendung von (1.3) auf den Quotientenraum E_U von E nach dem Nullraum einer Nullumgebung U in E, auf der die x'_n gleichmäßig beschränkt sind und die einer der Def. 1 genügenden Nullumgebungsbasis angehört. E_U wird in üblicher Weise mittels des auf den Restklassen (mod des Nullraumes von U) konstanten Wertes der Eichfunktion von U normiert; das kanonische Bild von K ist ein normaler Kegel in E_U.] Desgleichen folgt aus der strikten BZ-Eigenschaft (Def. 2', HV p. 128) von H $y_n = w_n - z_n$,

---

[13]) Wir setzen für den Rest dieses Abschnittes E, F als reelle Räume voraus. Die folgenden Überlegungen lassen sich leicht auf den Fall übertragen, daß E, F bzw. H, K komplexe Erweiterungen reeller Räume bzw. Kegel sind. Vgl. (6.5) und die anschließenden Bemerkungen.

[14]) Es sei E mit $\tau(E, E')$, F' mit irgendeiner (lokalkonvexen) Topologie versehen, die mit dem Dualsystem $\langle F, F' \rangle$ verträglich ist.

wo w_n, z_n ($n \in \mathbf{N}$) ein und derselben beschränkten abgeschlossenen Teilmenge von $\bar{H}$ angehören. Damit ist die Behauptung bewiesen.

Corollar. *Es sei E ein nuklearer tonnelierter Raum und K normal. F sei ein B-Raum, H ein abgeschlossener strikter BZ Kegel in F. Es gilt* $\mathfrak{L}(E, F) = \mathfrak{R} - \mathfrak{R}$.

Denn jede stetige lineare Abbildung von E in F ist nuklear ([7'] II, p. 35).

9. Stetigkeit und Konvergenz

Es seien E, F halbgeordnete lokalkonvexe Räume (über $\mathbf{R}$ oder $\mathbf{C}$), K, H die positiven Kegel dieser Halbordnungen in E, F. Die Frage der Stetigkeit einer positiven, d. h. $T(K) \subset H$ genügenden, linearen Abbildung von E in F ist eng mit den topologischen Eigenschaften der Kegel K und H verknüpft. Namioka ([19], p. 24) gibt eine interessante Bedingung für die Stetigkeit einer positiven linearen Abbildung, die wesentlich auf einem Kategorie-Argument beruht. Die folgenden Sätze, soweit sie sich auf die Stetigkeit positiver Abbildungen beziehen, sind Verallgemeinerungen bekannter[15]) Sätze über positive Linearformen und von Metrisierbarkeitsannahmen unabhängig.

In dem folgenden Satz braucht E nicht lokalkonvex zu sein.

(9.1). *Es sei E ein halbgeordneter topologischer linearer Raum, dessen positiver Kegel innere Punkte besitzt.* $F[\mathfrak{T}]$ *sei halbgeordnet durch H, und eine der Bedingungen sei erfüllt:*

a) $\mathfrak{T}$ *ist die Ordnungstopologie* $\mathfrak{T}_O$ *auf* F_0[16]).

b) H *ist normal für* $\mathfrak{T}$.

Jede positive lineare Abbildung T von E in F ist stetig.

Beweis. Es genügt, den Satz unter der Voraussetzung a) zu beweisen; denn ist H normal für $\mathfrak{T}$, so ist $\mathfrak{T}_O$ feiner als $\mathfrak{T}$ nach (4.4). Sei x_0 ein beliebiger innerer Punkt von K. Wir setzen $T x_0 = y_0$ und betrachten den reellen linearen Teilraum $L_{y_0} = \bigcup_{n=1}^{\infty} nM$ von F, $M = \{y: -y_0 \leq y \leq y_0\}$. Sei f die kanonische Einbettung von L_{y_0} in F. Ist nun V eine beliebige Nullumgebung in F, so enthält wegen der Stetigkeit von f [vgl. (4.4)] $\overset{-1}{f}(V)$ eine Nullumgebung $V_1 = \varepsilon M$, $\varepsilon > 0$, in L_{y_0}. Da x_0 innerer Punkt von K ist, ist $U = \{x \in E: -\varepsilon x_0 \leq x \leq \varepsilon x_0\}$ eine Nullumgebung in E, und wegen $T(K) \subset H$ ist $T(U) \subset f(V_1) \subset V$. Damit ist der Satz bewiesen.

Ist $T x_0 = 0$ für einen inneren Punkt von K, so folgt $T(U) = \{0\}$, da H ein echter Kegel in F ist. Wir erhalten daher noch

(9.2). *Es sei E ein durch K halbgeordneter topologischer linearer Raum, F ein halbgeordneter linearer Raum. Jede positive lineare Abbildung von E in F, die auf einem inneren Punkt von K verschwindet, verschwindet identisch*[17]).

---

[15]) bzw. in HV bewiesener.

[16]) F_0 ist der zu F assoziierte reelle Raum. Ist F durch H nicht regulär halbgeordnet, so ist $\mathfrak{T}_O$ nicht notwendig Hausdorffsch. (9.1) ist auch eine Verallgemeinerung von [9'], Cor. 5.4.

[17]) Es ist also $T = 0$, wenn $T \geq 0$ ist und auf einem Punkt von K verschwindet, bezüglich dessen K radial (*ausgeglichen*, [14]) ist.

Ferner ergibt (9.1) (vgl. BOURBAKI [7], p. 75) das

Corollar. *Ist E ein topologischer linearer Raum, K ein Kegel in E mit inneren Punkten, so ist jede auf K schwach positive (und erst recht jede auf K stark positive) Linearform stetig*[18].

Eine Verallgemeinerung von (2.8) ist

(9.3). *Es sei E ein bornologischer lokalkonvexer Raum, halbgeordnet durch einen folgenvollständigen strikten BZ-Kegel K. Ist H in $F[\mathfrak{T}]$ normal, so ist jede positive lineare Abbildung von E in F stetig.*

Beweis. Sei $T \geqq 0$. Wäre T nicht stetig, so wäre $\{T x_n\}_{n\in N}$ nicht beschränkt in F für eine beschränkte Folge $\{x_n\}$ in E, $x_n \in K$. Da H normal ist, erzeugen die reellen Halbnormen $y \to |\mathsf{R}\langle y, y'\rangle|$, $y' \in H'$, die schwache Topologie auf F (denn es gilt $F' = H' - H'$). Daher ist die Menge $\{\mathsf{R}\langle T x_n, y'\rangle\}_{n\in N}$ nicht beschränkt für ein $y' \in H'$; wir können annehmen: $\mathsf{R}\langle T x_n, y'\rangle > n$. Die Folge $z_n = \sum_{\nu=1}^{n} \frac{1}{\nu^2} x_\nu$ ist monoton und offenbar konvergent gegen ein $z \in K$. Nun ist einerseits $\mathsf{R}\langle T z_n, y'\rangle \leqq \mathsf{R}\langle T z, y'\rangle$ wegen $T \geqq 0$, andererseits $\mathsf{R}\langle T z_n, y'\rangle > > \sum_{\nu=1}^{n} \frac{1}{\nu}$ für $n \in N$. Dieser Widerspruch beweist die Behauptung.

Wie zu (2.8) ergibt sich als

Corollar. *Es sei $E = \lim\limits_{\to} E_\alpha[\mathfrak{T}_\alpha]$ induktiver Limes von B-Räumen, $K \cap E_\alpha = K_\alpha$ abgeschlossen und strikter BZ-Kegel in E_α. Ist H normal in F, so ist jede positive lineare Abbildung von E in F stetig.*

Denn T ist stetig von E in F, genau wenn die Einschränkung von T auf jedes E_α $\mathfrak{T}_\alpha$-stetig ist. Die Behauptung folgt jetzt durch Anwendung von (9.3) auf jedes E_α.

Die bisherigen Aussagen in diesem Abschnitt bezogen sich auf die Stetigkeit einer einzelnen linearen Abbildung von E in F. Wir beweisen jetzt einen zu (7.3) analogen Konvergenzsatz.

(9.4). *Es sei E ein halbgeordneter tonnelierter Raum, dessen positiver Kegel K E erzeugt: $E = K - K$. F sei ein halbreflexiver Raum, halbgeordnet durch einen normalen Kegel H. Schließlich sei $\mathfrak{H}$ eine* (für die durch den Kegel $\mathfrak{K}$ positiver linearer Abbildungen in $\mathfrak{L}(E, F)$ erzeugte Ordnungsstruktur) *aufsteigend (bzw. absteigend) gerichtete Teilmenge von $\mathfrak{L}(E, F)$, die einer der Bedingungen genügt:*

a) *$\mathfrak{H}$ ist majorisiert (bzw. minorisiert).*

b) *$\mathfrak{H}$ ist beschränkt* (für die Topologie der einfachen Konvergenz).

Dann konvergiert der Filter der Enden von $\mathfrak{H}$ für die Topologie der präkompakten Konvergenz gegen ein $T_0 \in \mathfrak{L}(E, F)$, und es gilt $T_0 = \sup \mathfrak{H}$ (bzw. $T_0 = \inf \mathfrak{H}$) für die durch $\overline{\mathfrak{K}}$ erzeugte Halbordnung in $\mathfrak{L}$[19]*.*

Beweis. Wir zeigen zunächst, daß aus der Existenz einer oberen (bzw. unteren, wenn $\mathfrak{H}$ absteigend gerichtet ist) Schranke die Beschränktheit von $\mathfrak{F}(\mathfrak{H})$ für die Topologie der einfachen Konvergenz folgt. Es sei

[18]) Def. 7 dehnt sich unmittelbar auf das Paar $\langle E, E'\rangle$ aus, auch wenn E' nicht total im algebraischen Dual E^* von E ist.

[19]) $\overline{\mathfrak{K}}$ im Sinne der Topologie einfacher Konvergenz. Vgl. (8.1).

$\mathfrak{H}_1 = \{T \in \mathfrak{H}: T \geqq T_1\}$, $T_1 \in \mathfrak{H}$, ein Ende von $\mathfrak{H}$. Ist S eine obere Schranke von $\mathfrak{H}$, so gilt $0 \leqq T - T_1 \leqq S - T_1$ für alle $T \in \mathfrak{H}_1$. Weiter ist K wegen $E = K - K$ nach (1.4) $\mathfrak{S}$-Kegel für die gesättigte Hülle des Systems aller endlichen Teilmengen von E, folglich nach (8.3) $\mathfrak{K}$ normal für die Topologie der einfachen Konvergenz auf $\mathfrak{L}$. Daher ist nach dem zweiten Corollar von (1.1) $\mathfrak{H}_1$ beschränkt, d. h. $\mathfrak{F}(\mathfrak{H})$ ein beschränkter Filter. (Entsprechend für nach unten ordnungsbeschränktes, absteigendes $\mathfrak{H}$.) Aus der Normalität von $\mathfrak{K}$ folgt zugleich, daß $\mathfrak{L}$ durch $\mathfrak{K}$ (und $\bar{\mathfrak{K}}$) regulär halbgeordnet ist.

Es ist nun leicht zu sehen, daß $\mathfrak{F}(\mathfrak{H})$ einfach gegen eine lineare Abbildung T_0 von E in F konvergiert. Denn für jedes feste $x \in K$ ist die Menge $\{Tx\}_{T \in \mathfrak{H}}$ gerichtet in F; hieraus und aus der Beschränktheit von $\mathfrak{F}(\mathfrak{H})$ folgt nach (7.3), daß der Filter der Enden dieser Menge gegen ein $y \in F$ konvergiert. Wegen $E = K - K$ konvergiert $\mathfrak{F}(\mathfrak{H})$ für jedes $x \in E$; der Limes ist eine lineare Abbildung $x \to y$ von E in F, die wir mit T_0 bezeichnen. Nach dem Banach-Steinhaus-Theorem ([8], p. 27) ist T_0 stetig und $\mathfrak{F}(\mathfrak{H})$ konvergiert gegen T_0 für die Topologie der präkompakten Konvergenz. $T_0 = \sup \mathfrak{H}$ (für die $\bar{\mathfrak{K}}$-Halbordnung) folgt aus (7.1).

Corollar. *Unter den Voraussetzungen von (9.4) sei für eine gesättigte Klasse $\mathfrak{S}$ beschränkter Teilmengen von E K $\mathfrak{S}$-Kegel, und $\mathfrak{L}(E, F)$ besitze für die $\mathfrak{S}$-Topologie denselben Dual wie für die Topologie präkompakter Konvergenz. Dann konvergiert $\mathfrak{F}(\mathfrak{H})$ für die $\mathfrak{S}$-Topologie.*

Beweis. Sei $\mathfrak{L}'$ der Dual von $\mathfrak{L}$ für die Topologie präkompakter Konvergenz. Nach (9.4) konvergiert $\mathfrak{F}(\mathfrak{H})$ für $\sigma(\mathfrak{L}, \mathfrak{L}')$; andererseits ist $\mathfrak{K}$ normal nach (8.3). Daher folgt die Behauptung aus (7.2).

Nehmen wir in (9.4) für F den Skalarkörper ($\mathbf{R}$ oder $\mathbf{C}$), über dem E definiert ist, und für H den Kegel $\mathbf{R}^+$ nichtnegativer reeller Zahlen (bzw. $\mathbf{R}^+$ oder $\mathbf{R}^+ + i\mathbf{R}^+$ für komplexes E), so erhalten wir das

Corollar. *Sei E ein durch K halbgeordneter, tonnelierter Raum, und $E = K - K$. Jede aufsteigend quasi-monotone Folge von Linearformen, die beschränkt oder majorisiert ist, konvergiert für die Mackeysche Topologie $\tau(E', E)$.*

Wir schreiben noch eine spezialisierte Form von (9.4) für stetige Endomorphismen in reflexiven Räumen auf.

(9.5). *Es sei E ein reflexiver lokalkonvexer Raum, halbgeordnet durch einen normalen BZ-Kegel K. Jede beschränkte, quasi-monotone Folge $\{T_n\}$ stetiger Endomorphismen von E konvergiert gegen einen stetigen Endomorphismus T_0 gleichmäßig auf jeder kompakten Teilmenge von E.*

Beweis. Wir bemerken zunächst, daß die Voraussetzung über die T_n ungeändert bleibt, wenn wir zu der durch $\bar{K}$ erzeugten (feineren) Ordnungsstruktur übergehen. $\bar{K}$ ist wieder normal, und nach (2.5) gilt $E = \bar{K} - \bar{K}$. Mit $E = F$ sieht man jetzt, daß die Voraussetzungen von (9.4) erfüllt sind, womit die Behauptung bewiesen ist.

Ist in (9.5) E insbesondere ein Montelscher Raum, so gilt also $T_n \to T_0$ gleichmäßig auf allen beschränkten Teilmengen von E.

10. Der Eigenwertsatz für positiv-kompakte lineare Abbildungen

Ein Endomorphismus T eines linearen topologischen Raumes E heißt kompakt, wenn $T(U)$ relativ kompakt ist für eine geeignete Nullumgebung U in E. Für kompakte Endomorphismen ist die bekannte Spektraltheorie in Banachräumen nahezu vollständig auf lokalkonvexe (Leray [11′], Altman [1′]) und allgemeinere topologische lineare Räume (Williamson [15′]) übertragen worden. Ein großer Teil der entsprechenden (weniger reichhaltigen) Aussagen in B-Räumen läßt sich auch für beschränkte Endomorphismen (d. h. Abbildungen, für die das Bild einer Nullumgebung beschränkt ist) in lokalkonvexen Räumen aufrechterhalten (s. z. B. Grothendieck [6′], Verf. [20], [14′]). Bezüglich der an E zu stellenden Vollständigkeitsforderungen kommt man mit Folgenvollständigkeit oder Quasivollständigkeit[20]) aus. Für kompakte Endomorphismen sind indessen — soweit es sich um die Existenz von Eigenvektoren zu von Null verschiedenen Eigenwerten handelt — Vollständigkeitseigenschaften von E ohne Belang, da die Erweiterung von T auf die vollständige Hülle $\tilde{E}$ die letztere in E abbildet[21]). Wir beschränken uns im folgenden auf lokalkonvexe Räume, obwohl ein Teil der Resultate auch in allgemeineren topologischen linearen Räumen Gültigkeit besitzt. Die betrachteten Räume können reell oder komplex sein. Wenn vom Spektrum $\sigma(T)$ (d. h. allen komplexen Zahlen λ, für die $\lambda - T$ kein topologischer Automorphismus ist) die Rede ist, so ist dies stets bezüglich der vollständigen Hülle $\tilde{E}$ (bzw. ihrer komplexen Erweiterung, falls E reell ist) zu verstehen. Wir fassen zunächst die folgenden bekannten (l. c.) Tatsachen zusammen als

Lemma 1. *Sei E ein folgenvollständiger lokalkonvexer Raum. Das Spektrum jeder beschränkten linearen Abbildung $T \in \mathfrak{L}(E)$ ist nicht leer, kompakt und besteht für kompaktes T nur aus isolierten Eigenwerten endlicher Vielfachheit (mit möglicher Ausnahme der 0). $r = \sup\{|\lambda| : \lambda \in \sigma(T)\}$ heißt Spektralradius von T und die Reihe*

$$(1) \qquad (\lambda - T)^{-1} = \sum_{n=0}^{\infty} \lambda^{-(n+1)} T^n$$

konvergiert, falls $|\lambda| > r$ ist, für die Topologie beschränkter Konvergenz in $\mathfrak{L}(E)$[22]).

Es sei $U = \{x : p(x) \leq 1\}$ eine Nullumgebung in E, für die $T(U)$ beschränkt ist. Dann gilt

$$(2) \qquad p_\alpha(Tx) \leq C_\alpha p(x) \qquad (\alpha \in A, x \in E)$$

für ein beliebiges System $\{p_\alpha\}_{\alpha \in A}$ von Halbnormen, das die Topologie von E erzeugt. Mit $p(T^n) = \sup\{p(T^n x) : x \in U\}$ ergibt sich

$$(3) \qquad r = \lim_{n\to\infty} \sqrt[n]{p(T^n)}\ ^{23)}.$$

[20]) E heißt quasivollständig, wenn jede seiner abgeschlossenen beschränkten Teilmengen vollständig ist (für die durch den uniformen Raum E induzierte uniforme Struktur).

[21]) Anders für präkompaktes T, vgl. [12′].

[22]) bzw. $\mathfrak{L}(E_1)$, E_1 die komplexe Erweiterung von E.

[23]) Die Existenz des Limes folgt wie in B-Algebren.

Die Aussagen des Lemmas können beispielsweise gewonnen werden, indem man den Quotientenraum E/N (N der Nullraum von p) mit Hilfe des auf den Restklassen mod N konstanten Wertes von p normiert und die T in E/N zugeordnete Abbildung betrachtet, die mit T beschränkt bzw. kompakt ist.

Man kann aber auch ohne Rückgriff auf einen Quotientenraum zum Ziel kommen. Hierzu betrachten wir die Algebra $\mathfrak{A}$ aller Endomorphismen A von E, die die Nullumgebung $U = \{x: p(x) \leqq 1\}$ auf eine Teilmenge eines Vielfachen von U abbilden. Die durch

$$A \to p(A) = \sup_{x \in U} p(A\,x)$$

auf $\mathfrak{A}$ erzeugte Topologie ist nicht notwendig separiert. Ist $T \in \mathfrak{L}(E)$ und T auf U beschränkt, so ist $T \in \mathfrak{A}$ und das gleiche gilt von den Potenzen von T, $(\lambda - T)$ und (falls vorhanden) ihren (stetigen) Inversen. Ist nun $(\lambda_0 - T)^{-1} = R(\lambda_0, T)$ vorhanden und in $\mathfrak{L}(E)$ für ein $\lambda_0 \neq 0$[24]), so konvergiert die Reihe

$$R(\lambda, T) = \lambda^{-1}\left[I + \sum_{n=0}^{\infty} (\lambda_0 - \lambda)^n\, T\, R(\lambda_0, T)^{n+1}\right] \tag{4}$$

in der λ_0-Umgebung $|\lambda - \lambda_0| < p(R_{\lambda_0})^{-1}$, $\lambda \neq 0$ für die Topologie der beschränkten Konvergenz in $\mathfrak{L}(E)$. Dies folgt nach (2) aus der für den Koeffizienten von $(\lambda_0 - \lambda)^n$ in (4) gültigen Abschätzung

$$p_\alpha(T R_{\lambda_0}^{n+1} x) \leqq C_\alpha p(R_{\lambda_0}^{n+1} x) \leqq C_\alpha p(x)\, p(R_{\lambda_0})^{n+1} \qquad (\alpha \in A,\ x \in E)\,.$$

$R(\lambda, T)$ ist also eine lokal-holomorphe Funktion mit Werten in $\mathfrak{L}(E)$ bzw. $\mathfrak{L}(E_1)$ und es ist nicht schwer einzusehen, daß (4) die Inverse von $\lambda - T$ darstellt. Die Resolventenmenge von T ist also offen und enthält nach (1) das Äußere des Kreises $|\lambda| \leqq r$. Weiter ergibt sich, daß $R(\lambda, T)$ von einem regulären Punkt λ_0 aus längs jedes Weges analytisch fortgesetzt werden kann, auf dem $p(R_\lambda)$ beschränkt bleibt und der 0 nicht berührt[24]). Hieraus folgen die Aussagen von Lemma 1 für beschränkte Endomorphismen, während die zusätzlichen Eigenschaften für kompaktes T in bekannter Weise direkt aus der Kompaktheit von T erschlossen werden können. Insbesondere ergibt sich noch

Lemma 2. *λ ist Randpunkt von $\sigma(T)$ genau dann, wenn $p(R_\lambda)$ längs keines ganz in der Resolventenmenge verlaufenden, nach λ führenden Weges beschränkt bleibt.*

Es gilt weiter der folgende bemerkenswerte Satz.

Lemma 3. *Es sei E (lokalkonvex und) folgenvollständig, T beschränkt und $\lambda \neq 0$. Ist $\lambda - T$ ein algebraischer Automorphismus von E, so ist es auch ein topologischer Automorphismus. Die Spektren von T und T'* (als eines beschränkten Endomorphismus des starken Duals E') *stimmen überein.*

Für einen Beweis, der allerdings von der erwähnten Quotientenbildung abhängt, s. [20] Hilfssatz 2.1 und Fußnote 12 sowie [14'], Abschnitt 4.

Es sei $E[\mathfrak{T}]$ ein lokalkonvexer Raum, K ein (nicht notwendig echter) Kegel in E. Ist $\{U_\alpha\}_{\alpha \in A}$ eine aus absolutkonvexen Mengen bestehende Nullum-

---

[24]) Ist $E[\mathfrak{T}]$ nicht normierbar, so ist 0 stets in $\sigma(T)$. Im normierten Fall gibt die mit λ multiplizierte Gl. (4), falls 0 regulär ist, für $\lambda = 0$ die Inverse von T.

gebungsbasis in E und E_K die lineare Hülle von K, so bilden die absolutkonvexen Hüllen $\Gamma(U_\alpha \cap K)$ eine Nullumgebungsbasis für eine lokalkonvexe Topologie $\mathfrak{T}_0$ in E_K[25]. Die Eichfunktionen p_α der U_α in E stimmen mit den Eichfunktionen $\dot{p}_\alpha$ der $\Gamma(U_\alpha \cap K)$ auf K überein. $\mathfrak{T}_0$ ist feiner als und im allgemeinen verschieden von $\mathfrak{T}$ auf E_K.

(10.1). *$\mathfrak{T}_0$ stimmt auf E_K mit $\mathfrak{T}$ überein, wenn K in $E_K[\mathfrak{T}]$ innere Punkte besitzt. Ist $E[\mathfrak{T}]$ metrisierbar (bzw. normierbar) und K vollständig, so ist $E_K[\mathfrak{T}_0]$ ein F-Raum (bzw. B-Raum). Jede (für die durch K induzierte Ordnungsstruktur) monotone Folge in E_K ist gleichzeitig für $\mathfrak{T}$ und $\mathfrak{T}_0$ konvergent oder divergent.*

Beweis. Offenbar ist $\mathfrak{T}_0$ feiner als $\mathfrak{T}$ auf E_K. Besitzt K innere Punkte in $E_K[\mathfrak{T}]$, so ist jedes $\Gamma(U_\alpha \cap K)$ eine $\mathfrak{T}$-Nullumgebung. Jede für $\mathfrak{T}_0$ konvergente Folge konvergiert offenbar für $\mathfrak{T}$; die umgekehrte Aussage für monotone Folgen ergibt sich aus der Übereinstimmung von p_α mit $\dot{p}_\alpha$ $(\alpha \in A)$ auf K. Ist $\mathfrak{T}$ metrisierbar, so kann diese Topologie durch eine aufsteigende Folge $\{p_n\}$ von Halbnormen definiert werden; dann sind auch die $\mathfrak{T}_0$ definierenden $\dot{p}_n$ aufsteigend auf E_K. Es sei $\{z_n\}$ eine Cauchyfolge in $E_K[\mathfrak{T}_0]$; es gibt eine Teilfolge, die wir wieder mit $\{z_n\}$ bezeichnen und für welche $\dot{p}_n(z_n - z_{n-1}) < \frac{1}{n^2}$ $(n \in \mathrm{N})$ gilt[26]. Nach Definition von $\mathfrak{T}_0$ gibt es eine Darstellung $z_n - z_{n-1} = x_n - y_n$ mit $x_n, y_n \in K$ und $p_n(x_n) < \frac{1}{n^2}$, $p_n(y_n) < \frac{1}{n^2}$. Wir erhalten

$$z_n = \sum_{\nu=1}^{n} x_\nu - \sum_{\nu=1}^{n} y_\nu \qquad (n \in \mathrm{N})$$

und die monotonen Folgen $\left\{\sum_{\nu=1}^{n} x_\nu\right\}$ und $\left\{\sum_{\nu=1}^{n} y_\nu\right\}$ konvergieren, da K als $\mathfrak{T}$-vollständig vorausgesetzt wurde, gegen $x \in K$ bzw. $y \in K$ für $\mathfrak{T}$, folglich auch für $\mathfrak{T}_0$; daher konvergiert $\{z_n\}$, womit die Vollständigkeit von $E_K[\mathfrak{T}_0]$ gezeigt ist.

Definition 8. *Es sei $E[\mathfrak{T}]$ ein lokalkonvexer Raum, K ein Kegel in E. Ein Endomorphismus T von E_K, der K invariant läßt, heiße K-kompakt (bzw. K-beschränkt), wenn T auf K $\mathfrak{T}$-stetig ist und eine $\mathfrak{T}$-Nullumgebung in K auf eine relativ $\mathfrak{T}$-kompakte (bzw. $\mathfrak{T}$-beschränkte) Teilmenge von K abbildet. Als K-Spektralradius r_K von T bezeichnen wir den Spektralradius von T in $E_K[\mathfrak{T}_0]$*[27].

Die gegebene Definition des K-Spektralradius r_K von T ist sinnvoll, da jedes K-kompakte (bzw. K-beschränkte) T offenbar ein beschränkter Endomorphismus von $E_K[\mathfrak{T}_0]$ ist. Ist T ein stetiger Endomorphismus von $E_K[\mathfrak{T}]$ und K ein Kegel in E mit kompakter bzw. beschränkter Basis[28], so folgt aus $T(K) \subset K$ die K-Kompaktheit bzw. K-Beschränktheit von T. Auf diese Weise wird den möglichen Eigenschaften von K Rechnung getragen. Die Definition von $\mathfrak{T}_0$ ergibt nun, mit Benutzung von (3),

[25]) Wir benutzen das Symbol $\mathfrak{T}_0$ in diesem Abschnitt ausschließlich für die in der angegebenen Weise konstruierte Topologie auf E_K.

[26]) Man setze $z_0 = 0$.

[27]) bzw. in der vollständigen Hülle von $E_K[\mathfrak{T}_0]$.

[28]) K ist ein Kegel mit kompakter Basis (bzw. beschränkter Basis), wenn für eine Nullumgebung U $U \cap K$ kompakt (bzw. beschränkt) ist. Vgl. das auf (10.4) folgende Beispiel.

(10.2). *Es sei T K-beschränkt für die Nullumgebung $U = \{x: p(x) \leqq 1\}$ in $E[\mathfrak{T}]$. Es gilt*

$$r_K = \lim_{n\to\infty} \sqrt[n]{p(T^n)}$$

mit $p(T^n) = \sup\{p(T^n x): x \in U \cap K\}$, $n \in \mathbf{N}$.

Nach Lemma 1 ist das Spektrum einer K-beschränkten Abbildung (als eines beschränkten Endomorphismus von $E_K[\mathfrak{T}_0]$), das wir mit $\sigma_K(T)$ bezeichnen wollen, kompakt. Außerhalb des Spektrums existiert die Resolvente $R(\lambda, T)$ als stetiger Endomorphismus der vollständigen Hülle $\tilde{E}_K$ von $E_K[\mathfrak{T}_0]$. Bezüglich der gegebenen Topologie ist dies nicht allgemein richtig. Jedenfalls gilt aber

(10.3). *Es sei T K-beschränkt, K folgenvollständig (für $\mathfrak{T}$) und ϱ die größte reelle Zahl in $\sigma_K(T) \cup \{0\}$. Dann ist für $\lambda > \varrho$ die Resolvente $R(\lambda, T)$ von T in $\tilde{E}_K$ eine $\mathfrak{T}$-stetige Abbildung von K in K.*

Beweis. Wir haben $\varrho \leqq r_K$ und betrachten zunächst den Fall $\lambda > r_K$. Da T $\mathfrak{T}$-stetig ist auf K (Def. 8), gilt das gleiche für die Partialsummen von (1), und wegen (2) konvergiert (1) gleichmäßig auf $U \cap K$. Daher ist $R(\lambda, T)$ für $\lambda > r_K$ $\mathfrak{T}$-stetig auf K. Nach Lemma 2 ist nun $p(R_\lambda)$ auf jedem abgeschlossenen Intervall reeller Zahlen $\lambda > \varrho$ gleichmäßig beschränkt, und durch Fortsetzung von $R(\lambda, T)$ — falls $\varrho < r_K$ ist — mittels (4) kann in endlich vielen Schritten jedes $\lambda > \varrho$ erreicht werden. Aus der gleichmäßigen Konvergenz von (4) auf $U \cap K$ folgt die Behauptung.

Zum Beweis des Hauptsatzes (10.4) benötigen wir ein weiteres Lemma. Wir folgen hier, mit einer Reihe notwendiger Modifikationen (da z. B. das Prinzip gleichmäßiger Beschränktheit linearer Abbildungen nicht zur Verfügung steht), einem Gedankengang von Bonsall ([4′], p. 62—64).

Lemma 4. *Es sei T eine K-kompakte Abbildung, K ein echter abgeschlossener Kegel in $E[\mathfrak{T}]$. Der K-Spektralradius r_K ist im Spektrum $\sigma_K(T)$.*

Beweis. Wir benutzen folgenden Hilfssatz ([4′], Lemma 4.1). Jede nichtbeschränkte Folge $\{a_n\}$ nichtnegativer reeller Zahlen enthält eine Teilfolge $\{a_{n_k}\}$ mit den Eigenschaften:

1. $a_{n_k} > k$ $(k \in \mathbf{N})$,
2. $a_{n_k} > a_j$ $(j < n_k,\ k \in \mathbf{N})$.

(Beweis durch Induktion.) Ist nun $r_K = 0$, so ist nach Lemma 1 nichts zu beweisen. Wir nehmen also $r_K > 0$ an. Wäre r_K kein Element von $\sigma_K(T)$, so wäre für alle λ eines Intervalles $0 < r_K - \varepsilon < \lambda$, $\varepsilon > 0$, $R(\lambda, T)$ vorhanden und eine [für $\mathfrak{T}_0$, und nach (10.3) auch für $\mathfrak{T}$] stetige Abbildung von K in sich. Da $x = R_\lambda y$ für jedes $n \in \mathbf{N}$ der Beziehung

$$x = \lambda^{-1} y + \lambda^{-2} T y + \cdots + \lambda^{-n} T^{n-1} y + \lambda^{-n} T^n x$$

genügt, schließt man aus $y \in K$

$$Tx = -y + \lambda x \geqq \lambda^{-n} T^n y \qquad (n \in \mathbf{N}) \tag{5}$$

für die durch K erzeugte Halbordnung. Sei jetzt λ eine feste Zahl in $(r_K - \varepsilon, r_K)$.

Nach Definition von r_K, vgl. (10.2), gilt $\lim_{n\to\infty} \lambda^{-n} p(T^n) = +\infty$. Für jedes $n \in N$ wählen wir $y_n \in K$ mit $p(y_n) = 1$ derart, daß $p(T^n y_n) > \frac{1}{2} p(T^n)$ gilt. Wegen $\lambda^{-n} p(T^n y_n) \to +\infty$ gibt es nach dem Hilfssatz eine Folge $\{n_k\}_{k\in N}$ natürlicher Zahlen, so daß $\lambda^{-n_k} p(T^{n_k} y_{n_k}) > k$ und

$$(6) \qquad p(T^{n_k} y_{n_k}) > \lambda p(T^{n_k - 1} y_{n_k - 1})$$

für alle $k \in N$ zutrifft. Wir setzen $w_k = T^{n_k-1} y_{n_k}/p(T^{n_k} y_{n_k})$, haben also $p(T w_k) = 1$ und erhalten aus (5)

$$(7) \qquad \frac{T x_k}{\lambda^{-n_k} p(T^{n_k} y_{n_k})} \geqq T w_k$$

mit $x_k = R_\lambda y_{n_k}$. Zur Abschätzung von $p(w_k)$ finden wir

$$p(w_k) \leqq p(T^{n_k - 1})/p(T^{n_k} y_{n_k}) \leqq p(T^{n_k-1})/\lambda p(T^{n_k-1} y_{n_k - 1}) < 2\lambda^{-1}$$

wegen (6) und $p(T^n y_n) > \frac{1}{2} p(T^n)$. Infolge der K-Kompaktheit von T hat die Folge $\{T w_k\}$ einen Limeswert $u \in K$, für den $p(u) = 1$ gilt[29]). Andererseits ist $p(x_k) \leqq p(R_\lambda)$ für alle k, also $\{p(x_k)\}$ beschränkt, während der Nenner der linken Seite von (7) nach $+\infty$ strebt. Aus (2) folgt jetzt, daß die linke Seite von (7) für $k \to \infty$ gegen 0 konvergiert. Wegen der Abgeschlossenheit von K hat man also $u \leqq 0$, was wegen der Echtheit von K im Widerspruch zu $p(u) = 1$ steht. Damit ist Lemma 3 bewiesen.

Wir kommen zum Hauptsatz in diesem Abschnitt.

(10.4) (Eigenwertsatz). *E sei ein lokalkonvexer Raum, K ein abgeschlossener echter Kegel in E, T eine K-kompakte Abbildung* (Def. 8). *Ist $r_K > 0$, so ist r_K Eigenwert von T mit einem Eigenvektor $x \in K$.*

Beweis. Nach Lemma 4 ist r_K in $\sigma_K(T)$ enthalten; daher gilt nach Lemma 2 $p(R_\lambda) \to \infty$ für $\lambda \downarrow r_K$, wo nach Definition von $\mathfrak{T}_0$, vgl. (10.2),

$$p(R_\lambda) = \sup\{p(R_\lambda x) : x \in U \cap K\}$$

ist, U eine Nullumgebung in E mit relativ kompaktem $T(U \cap K)$, p die Eichfunktion von U. Es gibt also eine Folge $y_n \in K$ $(n \in N)$ mit $p(y_n) \to 0$ und $p(R_{\lambda_n} y_n) = 1$, $\{\lambda_n\}$ eine abnehmende, gegen $r_K > 0$ konvergente Folge. Die Elemente $x_n = \lambda_n R_{\lambda_n} y_n$ genügen den Gleichungen $\lambda_n x_n - T x_n = \lambda_n y_n$ $(n \in N)$. Wir setzen $w_n = x_n - y_n$ und haben $\lambda_n w_n = T x_n \in K$, folglich $w_n \in K$. w_n genügt der Gleichung $(\lambda_n - T) w_n = T y_n$. Aus $p(y_n) \to 0$ folgt nach (2) $T y_n \to 0$. Daher ergibt sich $(\lambda_n - T) w_n \to 0$. Hieraus folgt, daß $\{w_n\}$ relativ kompakt ist für $\mathfrak{T}$ (T ist K-kompakt); also ist $\{w_n\}$ beschränkt, $(r_K - \lambda_n) w_n \to 0$ und $(r_K - T) w_n \to 0$. Ist W die abgeschlossene Hülle von $\{w_n\}$, so ist also $0 \in (r_K - T) W$, denn W ist kompakt und $r_K - T$ stetig. Es gibt also einen Eigenvektor $x \in W \subset K$, und da x offenbar als Häufungswert[29]) der Folge $\{w_n\}$ angenommen werden kann, ist der Satz bewiesen; denn für jeden Häufungswert x von $\{w_n\}$ gilt $p(x) = r_K$, also $x \neq 0$.

---

[29]) Wir verwenden für eine Folge und die zugehörige Menge das gleiche Symbol $\{w_n\}$. w heißt Häufungswert (oder Limeswert) von $\{w_n\}$, wenn für jede Umgebung U von w die Beziehung $w_n \in U$ für unendlich viele n gilt.

Zur Erläuterung der in diesem Abschnitt bisher diskutierten Fragen betrachten wir das folgende

Beispiel. Es sei $E[\mathfrak{T}]$ der Raum ω[30]), Produkt abzählbar unendlich vieler Exemplare von $\mathbf{R}$. Die Elemente $x \in \omega$ sind nichts anderes als beliebige Folgen $x = (x_1, x_2, \ldots)$ reeller Zahlen. Wir betrachten den Kegel

$$K = \{x\colon x_n \geqq x_{n+1} \geqq 0,\ n \in \mathbf{N}\}$$

aller nicht zunehmenden Folgen mit nichtnegativen Gliedern. $\mathfrak{T}$ wird auf ω durch das System von Halbnormen $x \to |x_n|$ $(n \in \mathbf{N})$ erzeugt. K ist vollständig in ω, und für die Nullumgebung $U_1 = \{x\colon |x_1| \leqq 1\}$ ist $U_1 \cap K$ kompakt. $E_K = K - K$ ist ein linearer Teilraum des Raumes (c) konvergenter Folgen. Da die Abbildung T:

$$x \to Tx = (x_2, x_3, \ldots)$$

ein stetiger Endomorphismus von ω ist und K invariant läßt, ist T K-kompakt. Der K-Spektralradius r_K von T bestimmt sich nach (10.2) sofort zu $r_K = 1$. Offenbar ist $e = (1, 1, \ldots)$ ein (und der einzige linear unabhängige) Eigenvektor von T in E_K zu r_K. (Das Spektrum von T in $E = \omega$ ist die ganze reelle Achse, bzw. in der komplexen Erweiterung von ω die ganze Ebene.) In $E_K[\mathfrak{T}_0]$ ist das Spektrum $\sigma_K(T)$ im Einheitskreis enthalten.

Die Topologie $\mathfrak{T}_0$ auf E_K ist feiner als die Topologie gleichmäßiger Konvergenz in allen Koordinaten. Zur Bestimmung von $E_K[\mathfrak{T}_0]'$ beachten wir: Da die Eichfunktionen der $\Gamma(U_n \cap K)$, $U_n = \{x\colon |x_n| \leqq 1\}$, auf K mit den $|x_n|$ übereinstimmen, ist K nach (1.1) normal für $\mathfrak{T}_0$ (denn K ist normal in ω). Nach (1.3) ist jede $\mathfrak{T}_0$-stetige Linearform Differenz zweier auf K positiver Linearformen. Wir setzen u_n gleich dem Vektor in ω, der an der Stelle n eine Eins und sonst nur Nullen hat. $z_n = \sum_{\nu=1}^{n} u_\nu$ ist in K $(n \in \mathbf{N})$; daher ist $u_n = z_n - z_{n-1}$ $(z_0 = 0)$ in E_K. Sei f eine auf K positive Linearform. [Nach (2.8) und (10.1) ist jede auf K positive Linearform stetig für $\mathfrak{T}_0$.] Mit $f(u_n) = c_n$ ist $f(z_n) = \sum_{\nu=1}^{n} c_\nu \geqq 0$; da $\{z_n\}$ beschränkt ist, besitzt $\sum_{\nu=1}^{\infty} c_\nu$ beschränkte nichtnegative Partialsummen. Die Reihe konvergiert im allgemeinen nicht; jedoch konvergiert $\sum_{\nu=1}^{\infty} c_\nu x_\nu$ für jede monotone Nullfolge, wie man mit Hilfe Abelscher partieller Summation sofort feststellt. Weiter ist leicht nachzuprüfen, daß $E_K[\mathfrak{T}_0] = [e] \oplus E_0$ direkte topologische Summe des linearen Teilraumes mit der Basis $\{e\}$ und des Teilraumes $E_0 = K_0 - K_0$ ist, wo $K_0 \subset K$ den Kegel nicht zunehmender Nullfolgen bezeichnet. Mit $f \in K'$, $c_0 = f(e)$ und $\xi = \lim_{\nu\to\infty} x_\nu$ ist nun

$$(*) \qquad f(x) = c_0 \xi + \sum_{\nu=1}^{\infty} c_\nu (x_\nu - \xi)$$

der Wert der Linearform f für jedes $x \in E_K$. Genügt umgekehrt eine Folge von Konstanten $(c_0; c_1, c_2, \ldots)$ den Bedingungen, daß $c_0 \geqq 0$ ist und $\sum_{\nu=1}^{\infty} c_\nu$ nicht-

[30]) Siehe z. B. [14] für zahlreiche Eigenschaften von ω.

negative beschränkte Partialsummen besitzt, so definiert (*) eine auf K positive (und folglich stetige, s. o.) Linearform auf $E_K[\mathfrak{T}_0]$. Wegen $E_K[\mathfrak{T}_0]' = K' - K'$ ist damit der duale Raum, der einem Teilraum von ω isomorph ist, vollständig bestimmt.

Es ist klar, daß die zu T in E'_K adjungierte Abbildung T' durch

$$f \to T'f \sim (c_0; 0, c_1, c_2, \ldots)$$

gegeben ist. $r_K = 1$ ist Eigenwert von T' mit (genau) einem (linear unabhängigen) Eigenvektor $(1;, 0, 0, \ldots)$. — Nach (10.1) ist $E_K[\mathfrak{T}_0]$ ein F-Raum. Da jedoch $\Gamma(U_1 \cap K)$ eine beschränkte konvexe Nullumgebung ist, ist $E_K[\mathfrak{T}_0]$ ein B-Raum. Man überzeugt sich leicht, daß die Norm $w \to \|w\| = \inf\{|x_1| + |y_1| : x, y \in K, x - y = w\}$ die Topologie $\mathfrak{T}_0$ auf E_K erzeugt.

In dem vorangehenden Beispiel war der K-Spektralradius r_K zugleich Eigenwert von T' in $E_K[\mathfrak{T}_0]'$. Dies wird man nicht immer erwarten können, da die K-Kompaktheit von T im allgemeinen nicht die K'-Kompaktheit von T' (für irgendeine geeignete Topologie auf $E_K[\mathfrak{T}_0]'$) nach sich zieht. Es gilt aber jedenfalls die folgende weitgehende Verallgemeinerung eines bekannten Satzes von Krein-Rutman [17]:

(10.5). *Sei E ein lokalkonvexer Raum, K ein Kegel in E, der echt, abgeschlossen und total ist. Ist T ein kompakter Endomorphismus von E mit $T(K) \subset K$ und positivem Spektralradius r, so ist r Eigenwert von T und T' mit je einem Eigenvektor $x \in K$ bzw. $x' \in K'$.*

Beweis. Das Resultat für T folgt aus (10.4), wenn wir zeigen, daß $r_K = r$ ist. Dies ergibt sich wie folgt: Nach (3) und (10.2) ist $r_K \leq r$. Sei $\tilde{E}$ bzw. $\tilde{E}_K$ die vollständige Hülle von $E[\mathfrak{T}]$ bzw. von $E_K[\mathfrak{T}_0]$ (wir nehmen E ohne Beschränkung der Allgemeinheit als komplex an). Da K in E total ist, ist $\tilde{E}_K$ in $\tilde{E}$ dicht [$\mathfrak{T}_0$ ist feiner als $\mathfrak{T}$ auf E_K, vgl. den (10.1) vorangehenden Absatz]. Wäre nun $r_K < r$, so wäre jedes λ mit $|\lambda| = r$ ein Punkt der Resolventenmenge von T in $\tilde{E}_K$, d. h. $\lambda - T$ ein $\mathfrak{T}_0$-Automorphismus von $\tilde{E}_K$ und folglich $(\lambda - T)\tilde{E}$ dicht in $\tilde{E}$, woraus die Eineindeutigkeit von $\lambda - T'$ in E' folgt. Da $|\lambda| = r$ einen Punkt $\lambda_0 \in \sigma(T)$ enthält und λ_0 wegen der Kompaktheit von T Eigenwert von T und T' ist[31]), kann $\lambda_0 - T'$ nicht eineindeutig sein; hieraus folgt $r_K = r$.

Die zu T adjungierte Abbildung T' in E' ist kompakt für die Topologie $\mathfrak{T}'_c$ der gleichmäßigen Konvergenz auf den kompakten Teilmengen von E (Köthe [10'], Satz 1), und der zu K konjugierte Kegel K' ist echt (da K total ist) und abgeschlossen für $\mathfrak{T}'_c$. Daher ist T' K'-kompakt (für $\mathfrak{T}'_c$) in E' (Def. 8); wie eben beweist man $r_{K'} = r'$[32]). Es ist also nach (10.4) r' Eigenwert von T' (mit einem Eigenvektor in K'), und wegen $r' = r$ ist der Satz bewiesen.

[31]) Nach Lemma 3 stimmen die Spektren $\sigma(T)$ von T in $\tilde{E}$ und $\sigma(T')$ von T' in E' (für die starke Topologie) überein. Da T' kompakt ist für $\mathfrak{T}'_c$ (s. den zweiten Teil des Beweises), ist jedes $0 \neq \lambda \in \sigma(T')$ Eigenwert.

[32]) Ist T kompakt, so stimmen die Spektren von T' bezüglich der Topologien $\mathfrak{T}'_c$ und $\mathfrak{T}'_b = \beta(E', E)$ auf E' überein.

Bemerkung. Überraschend an (10.5) ist nicht etwa, daß T und T' einen Eigenwert vom Betrag r besitzen (dies folgt schon aus Lemma 1), sondern daß für kompaktes T die Voraussetzung über K bereits die Relation $r \in \sigma(T)$ nach sich zieht. Vgl. hierzu Lemma 4 und (10.7).

Wir stellen nun eine Bedingung für die schwache K'-Kompaktheit von T' auf.

(10.6). *Sei E ein lokalkonvexer Raum*[9]*), K ein echter Kegel in E mit inneren Punkten. Ist T ein stetiger Endomorphismus von E, der K invariant läßt, so ist T' K'-kompakt für die schwache Topologie $\sigma(E', E)$.*

Beweis. Nach (7.4) ist K' ein Kegel in E' mit schwachkompakter Basis[28]). Da T' schwach stetig ist, ist T' K'-kompakt für die schwache Topologie (Def. 8).

Zu schwächeren Spektralaussagen als (10.4) oder (10.5) gelangt man, wenn man Kompaktheits- durch Beschränktheitseigenschaften von T ersetzt, dafür aber K als normalen Kegel voraussetzt.

(10.7). *Sei E lokalkonvex, K ein normaler Kegel in E. Ist T K-beschränkt* (Def. 8), *so ist $r_K \in \sigma_K(T)$.*

Beweis. Ist $r_K = 0$, so ist nach Lemma 1 nichts zu beweisen. Es sei also $r_K > 0$. Wäre $r_K \notin \sigma_K(T)$, so wäre die Resolvente $R(\lambda, T)$ von T in $\tilde{E}_K[\mathfrak{T}_0]$ für ein $\lambda < r_K$ vorhanden und $R_\lambda(\bar{K}) \subset \bar{K}$. Wie in Lemma 4 erhält man mit $y \in K$, $x = R_\lambda(y)$

$$Tx \geqq \lambda^{-n} T^n y \geqq 0 \qquad (n \in \mathrm{N}) \tag{5}$$

für die durch $\bar{K}$ in $\tilde{E}_K$ erzeugte Ordnungsstruktur. Nach Definition von r_K, vgl. (10.2), $\lambda^{-n} p(T^n y_n) \to \infty$ $(n \to \infty)$ für eine Folge $\{y_n\} \subset K$ mit $p(y_n) = 1$. Für $x_n = R_\lambda y_n$ ist wegen $p(x_n) \leqq p(R_\lambda)$ $\{p(x_n)\}$ beschränkt, daher ist $\{T x_n\}$ beschränkt nach (2). (5) ergibt daher einen Widerspruch zu (1.1), zweites Corollar.

Corollar. *Es sei $E[\mathfrak{T}]$ ein bornologischer Raum, K ein normaler BZ-Kegel in E. Ist T ein beschränkter Endomorphismus von E, der K invariant läßt, so ist der Spektralradius r im Spektrum von T.*

Beweis. Aus der Definition von $\mathfrak{T}_0$ folgt, daß die beschränkten Teilmengen von K für $\mathfrak{T}_0$ und $\mathfrak{T}$ stets dieselben sind. Unter den gegenwärtigen Voraussetzungen sind daher $\mathfrak{T}_0$ und $\mathfrak{T}$ auf E_K identisch. Daher ist $\sigma(T) = \sigma_K(T)$, $r = r_K$ und folglich, da K normal ist, $r \in \sigma(T)$ nach (10.7).

Als weiteres Corollar erhalten wir einen Satz von Bonsall ([3], th. 1):

Corollar. *Es sei E normiert, K ein normaler Kegel in E mit inneren Punkten. T sei ein Endomorphismus von E mit $T(K) \subset K$ und positivem Spektralradius r. r ist Eigenwert von T' mit einem Eigenvektor $x' \in K'$.*

Beweis. E ist bornologisch und K BZ-Kegel in E nach (2.6). T ist stetig nach (9.1), also beschränkt. Daher ist $r \in \sigma(T)$ nach dem vorangehenden Corollar. Aus der Normalität von K folgt nach (1.3) bzw. (6.1) $E' = K' - K'$, so daß $E'_{K'} = E'$ ist. Nach (10.6) ist T' K'-kompakt für $\sigma(E', E)$. Wäre nun $r_{K'} < r' = r$ (bzw. $r_{K'} > r$), so wäre $r - T'$ ein algebraischer Automorphismus

von $E'_{K'}$ (bzw. $r_{K'} - T'$ ein algebraischer Automorphismus von E'), was nicht sein kann. Daher ist $r_{K'} = r$ und die Behauptung folgt aus (10.4).

Infolge von (1.5), zweites Corollar, und (7.6) läßt sich dieser Satz nicht auf nichtnormierbare Räume verallgemeinern.

11. Einige Resultate über positive Transformationen in Banachschen und Hilbertschen Räumen

In diesem Abschnitt werden einige Ergebnisse zusammengestellt, deren gemeinsamer Kern darin besteht, einen Zusammenhang zwischen gewissen Eigenschaften des Spektrums eines stetigen Endomorphismus T und dem Verhalten von T bezüglich eines konvexen Kegels in dem zugrundeliegenden (Banachschen oder Hilbertschen) Raum herzustellen. Wir beweisen zunächst einen Satz, der für beliebige Banachalgebren mit Einselement gilt.

Es sei $\mathfrak{A}$ eine Algebra über $\mathbf{R}$ oder $\mathbf{C}$. Eine Ordnungsstruktur auf $\mathfrak{A}$ ist mit der algebraischen Struktur verträglich, wenn die Teilmenge $\mathfrak{R} = \{a: a \geqq 0\}$ von $\mathfrak{A}$ ein Kegel[4]) in dem zu $\mathfrak{A}$ assoziierten Vektorraum ist, der das Einheitselement von $\mathfrak{A}$ (falls ein solches existiert) und mit je zwei Elementen a, b deren Produkt ab enthält. (Einen solchen Kegel nennen wir *multiplikativ invariant.*) Ein Kegel in einer normierten Algebra heißt normal, wenn er multiplikativ invariant und ein normaler Kegel des zu $\mathfrak{A}$ assoziierten normierten linearen Raumes ist (Def. 1).

(11.1). *Es sei $\mathfrak{A}$ eine Banachalgebra mit Einselement, $\mathfrak{R}$ ein normaler Kegel in $\mathfrak{A}$. Für jedes $a \in \mathfrak{R}$ ist der Spektralradius $r(a) = \lim_{n\to\infty} \|a^n\|^{1/n}$ ein Element des Spektrums von a*[33]).

Beweis. Es sei $a \in \mathfrak{R}$ fest, $R(\lambda, a)$ die Resolvente von a (s. z. B. [8']). Aus der Funktionalgleichung der Resolvente (l. c. p. 126) folgt unmittelbar, daß eine Zahl λ Randpunkt des Spektrums $\sigma(a)$ genau dann ist, wenn für jede in der Resolventenmenge $\varrho(a)$ gelegene Folge $\{\lambda_n\}$ $\lim_{\lambda_n\to\lambda} \|R(\lambda_n, a)\| = \infty$ gilt. Es sei jetzt r der Spektralradius von a. Wir nehmen an, es wäre $r \in \varrho(a)$. Mit $\lambda = |\lambda| e^{i\varphi}$ $(0 \leqq \varphi < 2\pi)$ haben wir für $|\lambda| > r$

$$R(\lambda, a) = \sum_{n=0}^{\infty} |\lambda|^{-(n+1)} e^{-i(n+1)\varphi} a^n.$$

Mit

$$e^{-i(n+1)\varphi} = \alpha_{1,n} - \alpha_{2,n} + i(\alpha_{3,n} - \alpha_{4,n}),$$

wo $\alpha_{1,n} = \sup\{0, \cos(n+1)\varphi\}$, $\alpha_{2,n} = \sup\{0, -\cos(n+1)\varphi\}$, $\alpha_{3,n} = \sup\{0, \sin \times \times (n+1)\varphi\}$ und $\alpha_{4,n} = \sup\{0, -\sin(n+1)\varphi\}$ gesetzt ist, haben wir

$$(*) \qquad R(\lambda, a) = R_1 - R_2 + i(R_3 - R_4)$$

mit

$$R_j(|\lambda|, \varphi, a) = \sum_{n=0}^{\infty} |\lambda|^{-(n+1)} \alpha_{j,n} a^n \qquad (j = 1, \ldots, 4).$$

[33]) Im Beweis des Satzes wird nur benötigt, daß $\mathfrak{R}$ in dem zu $\mathfrak{A}$ assoziierten B-Raum normal ist, das Einselement und mit jedem a auch sämtliche Potenzen $a^n (n \in \mathbf{N})$ enthält.

Für die durch $\overline{\mathfrak{R}}$ in $\mathfrak{A}$ induzierte Ordnungsstruktur gilt aber offenbar $0 \leqq R_j \leqq R(|\lambda|, a)$, so daß nach dem zweiten Corollar von (1.1) aus der Beschränktheit von $R(\lambda, a)$ für $\lambda > r$ die Beschränktheit aller $R_j (j = 1, \ldots, 4)$ im Gebiet $|\lambda| > r$ folgt. Da $R(\lambda, a)$ auf $|\lambda| = r$ mindestens eine Singularität besitzt, muß also $r \in \sigma(a)$ sein und der Satz ist bewiesen.

Wir erhalten als Corollar folgenden Satz, der auf Bonsall ([4], th. 1)[34]) zurückgeht. Vgl. auch (10.7) und [4'], th. 2.

Corollar. *Es sei E ein Banachraum, K ein normaler BZ-Kegel in E. Für jeden stetigen Endomorphismus von E mit $T(K) \subset K$ ist der Spektralradius r im Spektrum von T.*

Der Beweis folgt unmittelbar aus (8.3) und (11.1).

Umgekehrt gestattet die Annahme, daß ein stetiger Endomorphismus T eines B-Raumes E keinen echten Kegel in E invariant läßt, zwei bemerkenswerte Folgerungen bezüglich des Spektrums von T.

(11.2). *Es sei E ein B-Raum, T ein stetiger Endomorphismus von E, der keinen echten Kegel $\neq \{0\}$ invariant läßt. Das Spektrum $\sigma(T)$ von T enthält keine nichtnegative Zahl, und das Punktspektrum von T ist nicht leer*[35]).

Beweis. Zum Beweis der ersten Behauptung nehmen wir an, es gäbe eine nichtnegative Zahl $\varrho \in \sigma(T)$. Ist ϱ ein Element des Punktspektrums, so transformiert T den Kegel $K = \{\lambda x_0 : \lambda \geqq 0\}$ in sich, $x_0 \neq 0$ ein beliebiger Eigenvektor zu ϱ. Dieser Fall kann also nicht eintreten. Daher ist $\varrho - T$ eineindeutig und kann folglich nicht E auf sich abbilden. Es gibt daher eine reelle Linearform $x \to f(x)$, die auf $(\varrho - T)E$ verschwindet, aber nicht identisch Null ist. T bildet den echten Kegel $\{0\} \cup \{x : f(x) > 0\}$ in sich ab. — Zum Beweis der zweiten Behauptung wählen wir ein festes $x_0 \neq 0$ und betrachten den Kegel K, der aus allen endlichen Summen der Gestalt $\sum_{\nu=0}^{n} a_\nu T^\nu x$ besteht, $a_\nu \geqq 0$. Da T K in sich abbildet, ist K nach der Voraussetzung über T nicht echt. Es gibt also eine kleinste natürliche Zahl n, für die $\sum_{\nu=0}^{n} a_\nu T^\nu x_0 = 0$ gilt mit $a_n > 0$. Daher ist der von $\{x_0, T x_0, \ldots, T^{n-1} x_0\}$ aufgespannte, mindestens eindimensionale lineare Teilraum invariant unter T. Daraus folgt, daß das Punktspektrum von T nicht leer ist.

Beispiel. Es sei C der B-Raum komplexwertiger stetiger Funktionen $t \to x(t)$ auf dem Intervall $[0, 1]$. Das Punktspektrum der linearen Abbildung T: $x \to i \int_0^t x(\tau)\, d\tau$ ist bekanntlich leer. Daher gibt es einen echten Kegel K, der unter T invariant ist. Die Funktionenmenge $\left\{ i^n \int_0^t (t-\tau)^{n-1} x_0(\tau)\, d\tau : n \in \mathrm{N} \right\}$ spannt für jedes feste $0 \neq x_0 \in C$ einen solchen Kegel K auf.

[34]) l. c. wird die Behauptung unter der Voraussetzung bewiesen, daß K in E und K' in E' (für die starke Topologie) normal sei. Dann ist aber K'' normaler BZ-Kegel in E'' nach (2.2), und man wird auf den gegenwärtigen Fall zurückgeführt.

[35]) Der Satz (und Beweis) bleibt nach Abschnitt 10, Lemma 3, für beschränkte Endomorphismen in folgenvollständigen lokalkonvexen Räumen gültig.

Es sei X ein reeller Hilbertraum, $(x, y) \to \langle x, y\rangle$ das innere Produkt in X. Herkömmlicherweise nennt man einen symmetrischen Endomorphismus A von X positiv, wenn $\langle A x, x\rangle \geqq 0$ ist für jedes $x \in X$[26]). Da das Produkt zweier in diesem Sinne positiver Abbildungen im allgemeinen nicht positiv ist, kann es keinen Kegel K in X geben, so daß die positiven Abbildungen A genau diejenigen sind, die K invariant lassen. Mit anderen Worten: Die durch den (normalen) Kegel der positiven Endomorphismen in der (Banachschen) Operatorenalgebra induzierte Ordnungsstruktur läßt sich nicht durch einen Kegel $K \subset X$ im Sinne von Abschnitt 8 erzeugen. Wir haben jedoch folgenden Satz:

(11.3). *Sei X ein reeller Hilbertraum, $\{A_\iota\}_{\iota\in I}$ eine beliebige Menge paarweise vertauschbarer, (symmetrischer) positiver Operatoren auf X. Zu jedem $x_0 \in X$ existiert ein Kegel K, der x_0 enthält und die Eigenschaften besitzt:*

1° *$A_\iota(K) \subset K$ für jedes $\iota \in I$.*

2° *K ist abgeschlossen und normal.*

3° *K ist identisch mit seinem konjugierten Kegel K' (folglich ist K BZ-Kegel und $X = K - K$).*

Beweis. Es sei $\mathfrak{A}$ ein maximales System paarweise vertauschbarer (symmetrischer) positiver Operatoren, das die Menge $\{A_\iota : \iota \in I\}$ enthält. (Ein solches System existiert nach dem Satz von Zorn.) Ein $0 \neq x_0 \in X$ werde fest gewählt; K bezeichne die Menge aller Kegel $K \subset X$, die den Bedingungen genügen:

a) $x_0 \in K$ und K ist echt.

b) Aus $x \in K$, $y \in K$ und $A \in \mathfrak{A}$ folgt $\langle A x, y\rangle \geqq 0$.

K ist nicht leer, da K den Kegel $\{\lambda x_0 : \lambda \geqq 0\}$ enthält. Durch die Relation $K_1 \subset K_2$ ist K induktiv (halb-) geordnet. (Man prüft sofort nach, daß die Vereinigung einer total geordneten Teilmenge von K a) und b) erfüllt.) Nach dem Satz von Zorn gibt es daher ein maximales Element $K \in \mathsf{K}$. Wir zeigen, daß K den Bedingungen des Satzes genügt.

Es seien $x_1 \in K$, $A_1 \in \mathfrak{A}$ fest gewählt. Wir setzen $A_1 x_1 = z_1$. Nach Definition von $\mathfrak{A}$ gilt $\langle A z_1, y\rangle \geqq 0$ für alle $A \in \mathfrak{A}$ und $y \in K$. Der Kegel $K_1 = \{K + \lambda z_1\}_{\lambda \geqq 0}$ ist ein Element von K. Denn a) $x_0 \in K_1$, und wäre K_1 nicht echt, so wäre $x + \lambda z_1 = 0$ für ein $0 \neq x \in K$ und ein $\lambda > 0$, also $\langle x + \lambda z_1, x\rangle = \langle x, x\rangle + \lambda\langle z_1, x\rangle = 0$, was widerspruchsvoll ist (da $\langle z_1, x\rangle \geqq 0$ ist nach b)). Also ist K_1 echt, und wegen $\langle A z_1, y\rangle \geqq 0$ für alle $A \in \mathfrak{A}$ und $y \in K$ ist auch b) für K_1 erfüllt. Aus der Maximalität von K folgt daher $z_1 \in K$; da x_1, A_1 beliebig waren, folgt $A_\iota(K) \subset K$ für alle $\iota \in I$, womit die erste Aussage von (11.3) bewiesen ist.

Nun ist die identische Abbildung in $\mathfrak{A}$; daher folgt aus $x \in K$, $y \in K$ stets $\langle x, y\rangle \geqq 0$ und

$$\|x + y\|^2 = \langle x + y, x + y\rangle = \langle x, x\rangle + 2\langle x, y\rangle + \langle y, y\rangle \geqq \|y\|^2,$$

so daß K normal ist nach (1.1). Weiter genügt die abgeschlossene Hülle von K offenbar a) und b), daher ist K abgeschlossen. Wegen $\langle x, y\rangle \geqq 0$ für $x \in K$, $y \in K$ ist offenbar K ein Teilkegel des zu K konjugierten Kegels K' (Abschnitt 1) und nach dem eben benutzten Argument folgt $K = K'$ daraus, daß K maximal ist.

[26]) Wir beschränken uns auf stetige Operatoren. Ist X komplex, so ist jeder Operator mit $\langle A x, x\rangle \geqq 0$ $(x \in X)$ hermetisch.

Nach (1.3) ergibt sich jetzt $X = K - K$ und hieraus nach (2.6), daß K BZ-Kegel ist in X. Damit ist der Satz bewiesen.

Bemerkung. Ist X ein komplexer Hilbertraum, so existiert unter den Voraussetzungen von (11.3) ein Kegel $K \subset X$, der den Bedingungen 1° und 2° von (11.3) genügt.

Corollar. *Es sei X ein reeller Hilbertraum, $\{A_\iota\}_{\iota \in I}$ eine beliebige Menge paarweise vertauschbarer, kompakter symmetrischer und positiver Operatoren. Dann gibt es einen Kegel K, der den Bedingungen 1° bis 3° von (11.3) genügt und ein vollständiges System von zu positiven Eigenwerten gehörigen linear unabhängigen Eigenvektoren sämtlicher A_ι ($\iota \in I$) enthält.*

Beweis. Jedes A_ι gestattet bekanntlich eine Spektraldarstellung $A_\iota = \sum_{\nu=1}^{\infty} \lambda_{\iota,\nu} P_{\iota,\nu}$ mit eindimensionalen orthogonalen Projektoren $P_{\iota,\nu}$ ($\iota \in I$, $\nu \in N$). Da die A_ι paarweise vertauschbar sind, gilt das gleiche für die Gesamtheit dieser Projektoren. Daher existiert ein Kegel K, der den Aussagen von (11.3) genügt und für den insbesondere $P_{\iota,\nu}(K) \subset K$ gilt. Jedes $P_{\iota,\nu}$ ist kompakt und besitzt den Spektralradius 1, daher nach (10.5) einen Eigenvektor $x_{\iota,\nu} \in K$. Da $x_{\iota,\nu}$ einziger linear unabhängiger Eigenvektor von $P_{\iota,\nu}$ ist, ist die Behauptung bewiesen.

Literatur

(Die im Text verwendeten, nicht mit einem ′ versehenen Nummern beziehen sich auf das Literaturverzeichnis in [13′].)

[1′] Altman, M.: On linear functional equations in locally convex linear spaces. Stud. Math. 13, 194—207 (1953). — [2′] Amemiya, I.: A generalization of Riesz-Fischer's theorem. J. Math. Soc. Japan 5, 553—554 (1953). — [3′] Ando, T.: Positive linear operators in semi-ordered linear spaces. J. Fac. Science, Hokkaido University, Ser. I, XIII, 3 & 4, 214—228 (1957). — [4′] Bonsall, F.: Linear operators in complete positive cones. Proc. London Math. Soc. 3, VIII (29), 53—75 (1958). — [5′] Bourbaki, N.: Topologie générale, chap. X. Act. ind. sci. 1084. Paris 1949. — [6′] Grothendieck, A.: Leçons sur les espaces vectoriels topologiques. Sao Paulo 1954. — [7′] Grothendieck, A.: Produits tensoriels et espaces nucleaires. I, II. Memoirs of the Amer. Math. Soc. 16 (1955). — [8′] Hille-Phillips: Functional analysis and semi-groups. Amer. Math. Soc. Coll. Publ. XXXI (Rev. Ed.) (1957). — [9′] Kist, J.: Locally o-convex spaces. Duke Math. J. 25 (4), 569—582 (1958). — [10′] Köthe, G.: Zur Theorie der kompakten Operatoren in lokalkonvexen Räumen. Port. Math. 13 (3), 97—104 (1954). — [11′] Leray, J.: Valeurs propres et vecteurs propres d'un endomorphisme complètement continu d'un espace vectoriel à voisinages convexes. Acta Sci. Math. 12 B, 177—186 (1950). — [12′] Ringrose, R.: Precompact linear operators in locally convex spaces. Proc. Cambridge Phil. Soc. 53, 581—591 (1957). — [13′] Schaefer, H.: Halbgeordnete lokalkonvexe Vektorräume. Math. Ann. 135, 115—141 (1958). — [14′] Schaefer, H.: On the Fredholm alternative in locally convex linear spaces. Erscheint in Stud. Math. — [15′] Williamson, J.: Compact linear operators in linear topological spaces. J. London Math. Soc. 29, 149—156 (1954). — [16′] Yamamuro, S.: Monotone completeness of normed semi-ordered linear spaces. Pac. J. Math. 7, 1715—1725 (1957).

(Eingegangen am 16. April 1959)

Berichtigung

Zu „Über algebraische Integralgleichungen mit nichtnegativen Koeffizienten" von H. Schaefer, Math. Ann. 137, 385—391 (1959).

S. 385, Z. 4 v. u.: Anstelle von „der absolut größte Eigenwert" lies „ein absolut größter Eigenwert.

MAASS, H.
Math. Annalen 138, 287—315 (1959)

Über die räumliche Verteilung der Punkte in Gittern mit indefiniter Metrik

Von
HANS MAASS in Heidelberg

Einleitung

Während das Problem der räumlichen Verteilung der Punkte und Untergitter in einem euklidischen Gitter einer Behandlung zugänglich, wenngleich auch noch nicht in voller Allgemeinheit gelöst ist [4], zeigte sich bisher kein Ansatzpunkt, um analoge Fragestellungen für Gitter mit indefiniter Metrik in Angriff nehmen zu können. In dieser Hinsicht scheint sich, wie im folgenden dargetan werden soll, das Prinzip der Fourierentwicklung in der Siegelschen Theorie der indefiniten quadratischen Formen [8] in gewisser Weise als nützlich zu erweisen.

Zu einer vorgegebenen rationalen symmetrischen Matrix $S = S^{(m)}$ der Signatur $(n, m - n)$ mit $0 < n < m$ bilde man in bekannter Weise die Thetareihe

$$f(z; S, P) = \sum_{\mathfrak{g}} e^{\pi i R[\mathfrak{g}]}, \tag{1}$$

wobei über alle ganzen Spalten $\mathfrak{g}$ summiert wird und $R = xS + iyP$ mit reellen x, y sowie $z = x + iy$ gesetzt ist. P bezeichnet eine Majorante von S, d. h. eine Lösung von

$$PS^{-1}P = S, \quad P > 0. \tag{2}$$

Eine Parameterdarstellung für die P wird durch

$$P = 2K - S, \quad K = W^{-1}[X'S], \quad W = S[X] > 0, \quad X = X^{(m,n)} \tag{3}$$

geliefert. Die Automorphismen U von S, d. h. die reellen U mit $S[U] = S$ bilden den zu S gehörigen Majorantenraum $\mathfrak{P}(S)$ vermöge $P \to P[U^{-1}]$ auf sich ab und lassen die durch

$$ds^2 = \sigma(P^{-1}dP\,P^{-1}dP) \qquad (\sigma = \text{Spur}) \tag{4}$$

definierte Metrik von $\mathfrak{P}(S)$ invariant. dv bezeichne das zugehörige invariante Volumenelement in $\mathfrak{P}(S)$. Die ganz rationalen U in der Gruppe $\Omega(S)$ aller Automorphismen von S bilden die volle Einheitengruppe von S. $\Gamma(S)$ sei irgendeine ihrer Untergruppen von endlichem Index und $\mathfrak{F}(S)$ ein Fundamentalbereich bezüglich $\Gamma(S)$ in $\mathfrak{P}(S)$. Das Volumen $V(S)$ von $\mathfrak{F}(S)$ ist endlich, wenn wir den Fall ausschließen, daß zugleich $m = 2$ und $\sqrt{-|S|}$ rational ist, was im folgenden geschehen soll. Im ternären Fall $m = 3$, $n = 1$, der uns besonders beschäftigen wird, ist $\Gamma(S)$ stets eine in gewisser Weise ausgewählte echte Untergruppe der vollen Einheitengruppe von S.

Für jede Funktion $f(P)$, die wie $f(z;\, S,\, P)$ gegenüber den Substitutionen $P \to P[U]$ ($U \in \Gamma(S)$) invariant ist, kann formal der Integralmittelwert

$$(5) \qquad I\{f(P)\} = \frac{1}{V(S)} \int\limits_{\mathfrak{F}(S)} f(P)\, dv$$

gebildet werden. Durch eingehendes Studium der Mittelwerte

$$(6) \qquad g(z,\, S) = I\{f(z;\, S,\, P)\}$$

gelangt C. L. SIEGEL sodann zu seiner analytischen Theorie der indefiniten quadratischen Formen. Im Hinblick auf gewisse Untersuchungen im Bereich der positiv definiten quadratischen Formen war es nun naheliegend, allgemein automorphe Funktionen der Art $f(P)$ in eine Reihe nach invarianten Eigenfunktionen des Laplace-Beltramischen Operators Δ_S von $\mathfrak{P}(S)$ zu entwickeln. Wir denken uns eine solche Eigenfunktion als zweimal stetig differenzierbare Funktion von X dargestellt und haben demgemäß zu fordern

(7)
1. $u(UXV) = u(X)$ für $U \in \Gamma(S)$, $|V| \neq 0$, $S[X] > 0$,
2. $\Delta_S\, u(X) + \lambda u(X) = 0$ mit konstantem $\lambda (\geqq 0)$,
3. die quadratische Integrierbarkeit von $u(X)$ über $\mathfrak{F}(S)$.

Der Reihenentwicklung von $f(P)$ entspricht die Bildung der Fourierkoeffizienten $I\{f(P)\, u(X)\}$. Damit werden wir auf die Funktionen

$$(8) \qquad g(z;\, S,\, u) = I\{f(z;\, S,\, P)\, u(X)\}$$

geführt.

Hinsichtlich der Vollständigkeit des Ansatzes ist zu bemerken, daß das kontinuierliche Spektrum von Δ_S hier außer acht bleibt. Mit dem Auftreten eines solchen ist immer dann zu rechnen, wenn $S[\mathfrak{x}]$ eine Nullform ist, d. h. stets im Falle $m \geqq 5$. Die Existenz von $g(z;\, S,\, u)$ ist dann gewährleistet, wenn $f(z;\, S,\, P)$ über $\mathfrak{F}(S)$ quadratisch integrierbar ist. Auf Grund der Abschätzungen in [8], S. 29 ist festzustellen, daß $m - q - 3 > 0$ eine hinreichende Bedingung für die quadratische Integrierbarkeit von $f(z;\, S,\, P)$ über $\mathfrak{F}(S)$ ist. Dabei ist q eine gewisse ganze Zahl im Intervall $0 \leqq q \leqq \operatorname{Min}(n,\, m - n)$, die genau dann verschwindet, wenn $S[\mathfrak{x}]$ keine Nullform ist. Diese etwas summarische Betrachtung ergibt die Gültigkeit aller Betrachtungen unter der Voraussetzung $m > 3 + \operatorname{Min}(n,\, m - n)$ oder der Annahme, daß $S[\mathfrak{x}]$ keine Nullform ist. Im besonderen ist zu beachten, daß im Falle $m = 3$ ein kompakter Fundamentalbereich $\mathfrak{F}(S)$ genau dann existiert, wenn $S[\mathfrak{x}]$ nicht Nullform ist, worauf bereits in [2], S. 530—532 hingewiesen wurde. Da in den Abschätzungen in [8], S. 29 an Stelle der unendlichen Reihen $f(z;\, S,\, P)$ jeweils die zugehörigen Betragreihen genommen werden, so sind auch die in den folgenden Rechnungen erforderlichen Vertauschungen von Summationen und Integrationen zu rechtfertigen.

Die explizite Bestimmung der Fourierentwicklung von $g(z;\, S,\, u)$ ist im wesentlichen mit der Berechnung des Integrals

$$(9) \qquad \chi(y) = \chi(y;\, S,\, \mathfrak{g},\, u) = |W|^{\frac{n+1-m}{2}} \int\limits_{\substack{S[X] = W \\ X \in \mathfrak{F}_0(S,\, \mathfrak{g})}} e^{-2\pi y w} u(X) \frac{[dX]}{[dW]}$$

geleistet. Hierin ist $\mathfrak{g}$ eine ganze Spalte, $w = (S[X])^{-1}[X'S\mathfrak{g}]$, W eine positive Matrix und $\mathfrak{F}_0(S, \mathfrak{g})$ ein Fundamentalbereich in $S[X] > 0$ bezüglich der Gruppe der Substitutionen $X \to UX$ mit $U \in \Gamma(S, \mathfrak{g}) = \{U : U \in \Gamma(S), U\mathfrak{g} = \mathfrak{g}\}$. Der Quotient der Volumenelemente $[dX]$ und $[dW]$ ist im Sinne von SIEGEL [8] zu verstehen. Die Methode, mit der $\chi(y)$ in [8] im Falle $u = 1$ berechnet wurde, beruht auf geschickter Anwendung von Teilintegrationen und Integraltransformationen; sie läßt sich aber nicht unmittelbar auf den allgemeinen Fall anwenden. Hier ist entscheidend, daß mit Hilfe einer Greenschen Formel zunächst die Differentialgleichung

$$y^2 \chi''(y) + y\left(2\pi t y + \frac{m}{2}\right)\chi'(y) + \left(\pi t y n + \frac{\lambda}{4}\right)\chi(y) = 0 \tag{10}$$

mit $t = S[\mathfrak{g}]$ bewiesen werden kann. Man erhält sie, indem man die Differentiationen nach y unter dem Integralzeichen ausführt. Die Zusammenfassung aller Integrale ergibt einen Integranden, der, wie eine längere Rechnung zeigt, mit $\frac{1}{4}u(X)(\Delta_S + \lambda)e^{-2\pi y w}$ identisch ist. Verschiebt man nun mit Hilfe der Greenschen Formel den Operator Δ_S von $e^{-2\pi y w}$ auf $u(X)$, so wird das Integral unmittelbar in 0 übergeführt, womit (11) bewiesen ist. Es sei ν eine Lösung von $\nu(\nu + m - 2) + \lambda = 0$. Als Lösung von (10) ist dann $\chi(y)$ von der Gestalt

$$\chi(y) = \begin{cases} a_0 y^{\frac{\nu}{2}} W_0\left(y, \frac{m}{2} + \nu\right) + b_0 y^{\frac{\nu}{2}} & \text{für } t = 0\,, \\ a(\pi|t|y)^{\frac{\nu}{2}} e^{-\pi t y} W\left(\pi|t|y; \frac{n+\nu}{2}, \frac{m-n+\nu}{2}, \operatorname{sgn} t\right) & \text{für } t \neq 0\,, \end{cases} \tag{11}$$

wenn

$$W_0(y, \gamma) = \frac{y^{1-\gamma} - 1}{1 - \gamma} = \sum_{\alpha=1}^{\infty} \frac{(\log y)^\alpha}{\alpha!}(1-\gamma)^{\alpha-1} \tag{12}$$

und

$$W(y; \alpha, \beta, \varepsilon) = y^{-\frac{\alpha+\beta}{2}} W_{\frac{(\alpha-\beta)\varepsilon}{2}, \frac{\alpha+\beta-1}{2}}(2y) \qquad (\varepsilon^2 = 1) \tag{13}$$

gesetzt wird. $W_{\varkappa,\mu}(y)$ bezeichnet die Whittakersche Funktion.

Man sieht nun leicht, daß die Funktion

$$h(z; S, u) = (\pi y)^{-\frac{\nu}{2}} g(z; S, u) \tag{14}$$

eine Fourierentwicklung der folgenden Art gestattet:

$$\begin{aligned} h(z; S, u) = {} & \alpha_0(S, u)\, W_0\left(y, \frac{m}{2} + \nu\right) + \beta_0(S, u) + \\ & + \sum_{t \neq 0} \alpha_t(S, u)|t|^{\frac{m+\nu}{2} - 1} W\left(\pi|t|y; \frac{n+\nu}{2}, \frac{m-n+\nu}{2}, \operatorname{sgn} t\right) e^{\pi i x t}. \end{aligned} \tag{15}$$

Eine genauere Diskussion, die auf die Einführung der Maßzahlen in [8] Bezug nimmt, zeigt, daß $\alpha_t(S, u)$ von der Gestalt

$$\alpha_t(S, u) = \frac{2^{-\frac{m}{4}} \pi^{\frac{m}{2}} \|S\|^{-\frac{1}{2}} M(S, t, u)}{\Gamma\left(\frac{m + (2n - m)\varepsilon}{4}\right)\mu(S)} \qquad (\varepsilon = \operatorname{sgn} t) \tag{16}$$

ist, wobei $\mu(S)$ bis auf einen konstanten Faktor mit $V(S)$ übereinstimmt und

$$(17)\qquad M(S, t, u) = \sum_{\substack{\mathfrak{g} \\ S[\mathfrak{g}]=t}} \mu(S, \mathfrak{g}, u)$$

ist mit gewissen Gewichtsfunktionen $\mu(S, \mathfrak{g}, u)$, die in jedem Fall Integralmittelwerte von $u(X)$ über gewissen Bereichen sind. Summiert wird hier über ein volles System ganzer, bezüglich $\Gamma(S)$ paarweise nicht assoziierter Lösungen $\mathfrak{g}$ von $S[\mathfrak{g}] = t$. Genauere Aussagen sind nur im Fall $n = 1$ möglich. Dann ist der Majorantenraum mit dem $(m-1)$-dimensionalen hyperbolischen Raum identisch. Besonders einfach gestalten sich daher die Verhältnisse im ternären Fall $m = 3$, $n = 1$ in welchem

$$(18)\qquad \mu(S, \mathfrak{g}, u) = \begin{cases} \dfrac{\pi \|S\|^{-\frac{3}{2}} t^{-\frac{1}{2}}}{E(S,\mathfrak{g})} u(\mathfrak{g}) & \text{für} \quad t > 0, \\ \|S\|^{-\frac{3}{2}} |t|^{-\frac{1}{2}} l(S, \mathfrak{g})\, u(\mathfrak{g}) & \text{für} \quad t < 0 \end{cases}$$

gilt. Dabei ist $E(S, \mathfrak{g})$ die Ordnung von $\Gamma(S, \mathfrak{g})$ im Falle $t > 0$ und $l(S, \mathfrak{g})$ die Verschiebungslänge einer erzeugenden Substitution in $\Gamma(S, \mathfrak{g})$ sowie $u(\mathfrak{g})$ der Integralmittelwert von $u(\mathfrak{x})$ längs der hyperbolischen Geraden, die als Polare zu $\mathfrak{g}$ gekennzeichnet werden kann, sofern $t < 0$ ist. Hinsichtlich der ersten Koeffizienten in (15) wird nur gezeigt, daß

$$(19)\qquad \alpha_0(S, u) = 0\,, \quad \beta_0(S, u) = I\{u\}$$

ist, wenn $S[\mathfrak{x}]$ keine Nullform ist.

Aus der Transformationsformel

$$(20)\qquad f\left(-\frac{1}{z}; S, P\right) = \|S\|^{-\frac{1}{2}} (-iz)^{\frac{n}{2}} (i\bar{z})^{\frac{m-n}{2}} f(z; S^{-1}, P^{-1})\,,$$

die in der üblichen Weise zu beweisen ist, erhält man nach Multiplikation mit $u(X)$ und Bildung der Integralmittelwerte unmittelbar

$$(21)\qquad h\left(-\frac{1}{z}; S, u\right) = \|S\|^{-\frac{1}{2}} (-iz)^{\frac{n+\nu}{2}} (i\bar{z})^{\frac{m-n+\nu}{2}} h(z; S^{-1}, \tilde{u})$$

mit $\tilde{u}(X) = u(S^{-1}X)$. Die Entwicklung (15) setzt

$$(22)\qquad \left\{y^2\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) + \frac{m-2n}{2}\, i y \frac{\partial}{\partial x} + \left(\frac{m}{2} + \nu\right) y \frac{\partial}{\partial y}\right\} h(z; S, u) = 0$$

in Evidenz. Damit ist der Anschluß an die Theorie der Funktionen mit Dirichletscher Reihenentwicklung und Riemannscher Funktionalgleichung [6] gewonnen. Hier sei nur hervorgehoben, daß gewisse Linearkombinationen $\xi(s; S, u)$ und $\eta(s; S, u)$ von

$$(23)\qquad \begin{aligned} \varphi\left(s - \frac{\nu}{2}; S, u\right) &= \sum_{t>0} \alpha_t(S, u)\, t^{\frac{m+\nu}{2} - 1 - s}, \\ \psi\left(s - \frac{\nu}{2}; S, u\right) &= \sum_{t<0} \alpha_t(S, u)\, (-t)^{\frac{m+\nu}{2} - 1 - s} \end{aligned}$$

mit Gammafaktoren der Art

$$(24)\qquad \Gamma(s; \alpha, \beta) = \int_0^\infty W(y; \alpha, \beta, 1)\, y^{s-1} dy$$

als Koeffizienten meromorphe Funktionen von s sind. Um daraus den mero-

morphen Charakter von $\varphi(s; S, u)$ und $\psi(s; S, u)$ erschließen zu können, muß man wissen, daß sich die Nennerdeterminante

$$D(s; \alpha, \beta) = \begin{vmatrix} \Gamma(s; \alpha, \beta) & \Gamma(s; \beta, \alpha) \\ -\Gamma(s+1; \alpha, \beta) & \Gamma(s+1; \beta, \alpha) \end{vmatrix} \tag{25}$$

die bei der Auflösung nach $\varphi(s; S, u)$ und $\psi(s; S, u)$ erscheint, vernünftig verhält. Es ergibt sich die bemerkenswert einfache Formel

$$D(s; \alpha, \beta) = 2\,\Gamma(s)\,\Gamma(s+1-\alpha-\beta), \tag{26}$$

die alle erforderlichen Schlüsse gestattet. Dieser Sachverhalt ist auch noch für das allgemeine Theorem in [6], S. 256 von Bedeutung.

Die letzten Betrachtungen sind einem Gitterpunktproblem im ternären Fall $m = 3$, $n = 1$ gewidmet. Hier ist vorauszusetzen, daß $S[\mathfrak{x}]$ keine Nullform ist. Die Theorie der Eigenfunktionen $u(X)$ zu Δ_S gestaltet sich nun besonders einfach, weil das Spektrum des Operators Δ_S diskret ist und jeder Eigenwert λ endliche Vielfachheit hat. Wir denken uns den Fundamentalbereich $\mathfrak{F}(S)$ als Normalpolygon gewählt. $\mathfrak{B}$ bezeichne einen beliebigen, im Riemannschen Sinne meßbaren Teilbereich von $\mathfrak{F}(S)$. $\mathfrak{B}_0$ sei der entsprechende Bereich in $S[\mathfrak{x}] > 0$, der mit $\mathfrak{x}$ auch $r\mathfrak{x}$ $(r > 0)$ enthalte. Schließlich sei $\alpha_q(S, \mathfrak{B})$ die Anzahl der ganzen $\mathfrak{g} \in \mathfrak{B}_0$ mit $0 < S[\mathfrak{g}] \leq q$. Aus den analytischen Eigenschaften der Funktionen $\varphi(s; S, u)$ wird dann nach einem Verfahren, das ursprünglich auf Hecke zurückgeht, und das sich auch in [4] bewährt hat, die asymptotische Formel

$$\alpha_q(S, \mathfrak{B}) \sim \frac{2\pi}{3} \|S\|^{-\frac{1}{2}} V(\mathfrak{B})\, q^{\frac{3}{2}} \quad \text{für} \quad q \to \infty \tag{27}$$

gewonnen. Hierbei ist $V(\mathfrak{B})$ der hyperbolische Inhalt von $\mathfrak{B}$.

§ 1. Transformationen im Majorantenraum

Allgemein sei $E^{(k)}$ die k-reihige Einheitsmatrix und $C = C^{(m)}$ eine feste reelle Matrix mit der Eigenschaft

$$S[C] = S_0 = \begin{pmatrix} E^{(n)} & 0 \\ 0 & -E^{(m-n)} \end{pmatrix}. \tag{28}$$

Sodann wird mit reellem $Y = Y^{(n, m-n)}$ und nicht-singulärem $H = H^{(n)}$

$$X = C \begin{pmatrix} E \\ -Y' \end{pmatrix} H \tag{29}$$

gesetzt. Die über (3) hergestellte Beziehung $Y \to P$ bildet das Gebiet $E - YY' > 0$, wie bereits in [8] festgestellt wurde, umkehrbar eindeutig auf den Majorantenraum $\mathfrak{P}(S)$ ab. Die metrische Fundamentalform und das zugehörige Volumenelement stellen sich in der Gestalt

$$ds^2 = \sigma\{(E - YY')^{-1}\, dY (E - Y'Y)^{-1}\, dY'\} \tag{30}$$

und

$$dv = |E - YY'|^{-\frac{m}{2}}\, [dY] \tag{31}$$

dar. Die Isometrie $P \to P[U^{-1}]$ $(U \in \Omega(S))$ entspricht der Abbildung $X \to UX$ und induziert im Y-Raum die Transformation

$$Y \to (YC_0 + D_0)^{-1}(YA_0 + B_0)\,, \tag{32}$$

wobei

$$U_0 = C^{-1}UC = \begin{pmatrix} -D_0' & C_0' \\ B_0' & -A_0' \end{pmatrix}, \quad S_0[U_0] = S_0 \tag{33}$$

ist.

Der Fall $n = 1$ gestattet eine etwas ausführlichere Diskussion. Wir setzen nun

$$\begin{aligned} Y &= \mathfrak{y}' = (y_\beta) \qquad (\beta = 1, 2, \ldots, m-1)\,, \\ X &= \mathfrak{x} = (x_\alpha) \qquad (\alpha = 0, 1, \ldots, m-1)\,. \end{aligned} \tag{34}$$

Die Elemente von $\mathfrak{x}$ können zufolge $\mathfrak{x} = C\begin{pmatrix} 1 \\ -\mathfrak{y}\end{pmatrix} h\ (h \neq 0)$ als ein System homogener Koordinaten in $\mathfrak{P}(S)$ angesehen werden. $\mathfrak{P}(S)$ erscheint nun als Bild der offenen Kugel $y_1^2 + y_2^2 + \cdots + y_{m-1}^2 < 1$. Eine umkehrbar eindeutige Abbildung von $\mathfrak{P}(S)$ in den Halbraum $\xi_1 > 0$ vermittelt das System der Gleichungen

$$y_1 = \frac{\xi_1^2 + \cdots + \xi_{m-1}^2 - 1}{\xi_1^2 + \cdots + \xi_{m-1}^2 + 1}\,, \quad y_\alpha = \frac{2\xi_\alpha}{\xi_1^2 + \cdots + \xi_{m-1}^2 + 1} \quad (\alpha = 2, 3, \ldots, m-1) \tag{35}$$

und das dazu inverse System

$$\xi_1 = \frac{\sqrt{1 - y_1^2 - \cdots - y_{m-1}^2}}{1 - y_1}\,, \quad \xi_\alpha = \frac{y_\alpha}{1 - y_1} \quad (\alpha = 2, 3, \ldots, m-1)\,. \tag{36}$$

Für die metrische Fundamentalform ergibt sich nun nach einfacher Rechnung der Ausdruck

$$ds^2 = \xi_1^{-2}(d\xi_1^2 + d\xi_2^2 + \cdots + d\xi_{m-1}^2)\,. \tag{37}$$

$\mathfrak{P}(S)$ ist demnach mit dem $(m-1)$-dimensionalen hyperbolischen Raum identisch und der Laplace-Beltramische Operator lautet

$$\Delta_S = \xi_1^{m-1} \sum_{\alpha=1}^{m-1} \frac{\partial}{\partial \xi_\alpha}\left(\xi_1^{3-m} \frac{\partial}{\partial \xi_\alpha}\right). \tag{38}$$

Von besonderem Interesse ist eine geometrische Interpretation der Funktion $w = (S[X])^{-1}[X'S\mathfrak{g}]$, wobei $\mathfrak{g}$ eine von 0 verschiedene Spalte bezeichnet. Im vorliegenden Fall ist also, wenn $\mathfrak{q} = (q_\alpha)$ $(\alpha = 0, 1, \ldots, m-1)$ durch $\mathfrak{g} = C\mathfrak{q}$ eingeführt wird,

$$\begin{aligned} w &= \left(1 - \sum_{\alpha=1}^{m-1} y_\alpha^2\right)^{-1} \left(q_0 + \sum_{\alpha=1}^{m-1} q_\alpha y_\alpha\right)^2 \\ &= \frac{1}{4\xi_1^2}\left\{(q_0 + q_1)\sum_{\alpha=1}^{m-1} \xi_\alpha^2 + 2\sum_{\alpha=2}^{m-1} q_\alpha \xi_\alpha + q_0 - q_1\right\}^2. \end{aligned} \tag{39}$$

Wir setzen voraus, daß

$$S[\mathfrak{g}] = S_0[\mathfrak{q}] = t \tag{40}$$

von 0 verschieden ist. Entsprechend dem Vorzeichen von t läßt sich eine Isometrie $(\xi_\alpha) \to (\hat{\xi}_\alpha)$ angeben, so daß

$$w = \begin{cases} \frac{t}{4}\left(\frac{\hat{\xi}_1^2 + \hat{\xi}_2^2 + \cdots + \hat{\xi}_{m-1}^2 + 1}{\hat{\xi}_1}\right)^2 & \text{für } t > 0\,, \\ -t\left(\frac{\hat{\xi}_2}{\hat{\xi}_1}\right)^2 & \text{für } t < 0 \end{cases} \tag{41}$$

wird. Nun ist, wie leicht festgestellt werden kann,

$$\frac{\hat{\xi}_1^2 + \hat{\xi}_2^2 + \cdots + \hat{\xi}_{m-1}^2 + 1}{\hat{\xi}_1} = 2\operatorname{Cos}\varrho\,, \tag{42}$$

wenn ϱ den hyperbolischen Abstand vom Punkt $\hat{\xi}_1 = 1$, $\hat{\xi}_\alpha = 0$ $(\alpha > 1)$ bezeichnet und

$$\frac{\hat{\xi}_2}{\hat{\xi}_1} = \operatorname{Sin}\varrho\,, \tag{43}$$

wobei jetzt $|\varrho|$ den hyperbolischen Abstand von der hyperbolischen Ebene $\hat{\xi}_2 = 0$ bedeutet. Beachtet man noch, daß im Falle $t < 0$ die durch

$$(q_0 + q_1)\sum_{\alpha=1}^{m-1} \xi_\alpha^2 + 2\sum_{\alpha=1}^{m-1} q_\alpha \xi_\alpha + q_0 - q_1 = 0 \tag{44}$$

bestimmte hyperbolische Ebene mit dem in $S[\mathfrak{x}] > 0$ gelegenen Teil der Polaren zu $\mathfrak{g}$ bezüglich der Hyperfläche $S[\mathfrak{x}] = 0$ identisch ist, so kann auf Grund des Invarianzcharakters von ϱ leicht geschlossen werden, daß

$$w = \begin{cases} t\operatorname{Cos}^2\varrho & \text{für } t > 0\,, \\ -t\operatorname{Sin}^2\varrho & \text{für } t < 0 \end{cases} \tag{45}$$

gilt, wobei $\varrho = \varrho(\mathfrak{x}, \mathfrak{g})$ im Falle $t > 0$ den hyperbolischen Abstand der Punkte $\mathfrak{x}$ und $\mathfrak{g}$ und im Falle $t < 0$ vom Vorzeichen abgesehen den hyperbolischen Abstand des Punktes $\mathfrak{x}$ von der zu $\mathfrak{g}$ gehörigen Polaren bezeichnet.

Die Relationen zwischen den Automorphismen U von S und den Bewegungen des hyperbolischen Raumes lassen sich im ternären Fall $m = 3$, $n = 1$, auf den sich die letzten Betrachtungen dieses Paragraphen beziehen, auf Grund einer speziellen Parameterdarstellung besonders einfach beschreiben. Gemäß (33) genügt die Kenntnis der Automorphismen U_0 von S_0. Da nur die Abbildungen $P \to P[U^{-1}]$ interessieren, wir folglich zwischen U_0 und $-U_0$ nicht zu unterscheiden brauchen, genügt es, aus den Restklassen der Faktorgruppe $\Omega(S_0)/\{\pm E\}$ jeweils einen Repräsentanten auszuwählen. Das geschieht bekanntlich [2] durch die Parameterdarstellung

$$U_0 = \begin{pmatrix} \frac{1}{2}(a^2 + b^2 + c^2 + d^2) & \frac{1}{2}(-a^2 + b^2 - c^2 + d^2) & -ab - cd \\ \frac{1}{2}(-a^2 - b^2 + c^2 + d^2) & \frac{1}{2}(a^2 - b^2 - c^2 + d^2) & ab - cd \\ -ac - bd & ac - bd & ad + bc \end{pmatrix}, \tag{46}$$

wobei a, b, c, d reelle Zahlen mit der Determinante $ad - bc = \varepsilon\,(\varepsilon^2 = 1)$ sind. Die Abbildung $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \to U_0$ ist ein Homomorphismus mit dem Kern $\{\pm E\}$. Es sei nun $\Gamma(S)$ die Gruppe derjenigen Einheiten $U = CU_0C^{-1}$ von S, die

durch (46) mit $\varepsilon = 1$ geliefert werden. Das heißt $\Gamma(S)$ hat in der Gruppe aller Einheiten von S den Index 2 oder 4. Diese Auswahl von $\Gamma(S)$ entspricht der Beschränkung in der Bewegungsgruppe der hyperbolischen Ebene auf die winkeltreuen Abbildungen. Wird

(47) $$\xi = \xi_2\,, \quad \eta = \xi_1\,, \quad \tau = \xi + i\,\eta$$

gesetzt, so ist

(48) $$\tau \to \frac{a\tau + b}{c\tau + d}$$

die Darstellung der Bewegung, welche durch die Abbildung $P \to P[U^{-1}]$ $(U \in \Gamma(S))$ definiert wird.

Wir betrachten die Automorphismen U von S, die durch (46) geliefert werden und eine von 0 verschiedene Spalte $\mathfrak{g}$ festlassen: $U\mathfrak{g} = \mathfrak{g}$. Gleichwertig damit ist $U_0\mathfrak{q} = \mathfrak{q}$ mit $\mathfrak{g} = C\mathfrak{q}$. Ist U_0 vorgegeben, so existiert stets eine Lösung $\mathfrak{q} \neq 0$; denn es ist $|U_0 - E| = 0$ bei jeder Wahl der Parameter. Sind $A_{\alpha\beta}$ die algebraischen Komplemente zu den Elementen der Matrix $U_0 - E$, so ist

(49) $$(A_{\beta\alpha}) = \mathfrak{q}_0\mathfrak{q}_0' S_0 \quad \text{mit} \quad \mathfrak{q}_0' = (b - c, -b - c, a - d)\,.$$

Unter der Voraussetzung $U \neq E$ ist $\mathfrak{q}_0 \neq 0$, also $\mathfrak{q}$ durch U_0 bis auf einen skalaren Faktor eindeutig bestimmt:

(50) $$\mathfrak{q} = \gamma\mathfrak{q}_0,\ \gamma \neq 0\,.$$

Hieraus ist $t = S_0[\mathfrak{q}] = q_0^2 - q_1^2 - q_2^2 = \gamma^2(4 - (a + d)^2)$ zu schließen. Die Matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ ist also elliptisch, parabolisch oder hyperbolisch, je nachdem $t > 0$, $t = 0$ oder $t < 0$, d. h. $\mathfrak{g}$ ein elliptischer, parabolischer oder hyperbolischer Punkt ist. Die Fixpunkte der Substitution (48) sind durch $\mathfrak{q}$ eindeutig bestimmt. Es sind dies die Punkte $(-q_2 \pm \sqrt{-t})\,(q_0 + q_1)^{-1}$, die wir im folgenden mit τ_1, τ_2 bezeichnen werden. Es sei $\mathfrak{Im}\,\tau_1 > \mathfrak{Im}\,\tau_2$ bzw. $\tau_1 > \tau_2$ je nachdem $t > 0$ bzw. $t < 0$ ist. Ist $t < 0$, so schneidet die Polare zu $\mathfrak{g}$ den Kegelschnitt $S[\mathfrak{x}] = 0$ in den Punkten τ_1, τ_2. Normalkoordinaten ϑ, ϱ werden zu einem gegebenen nicht parabolischen Punkt $\mathfrak{q}$ durch folgende Gleichungen eingeführt:

(51) $$\hat{\tau} = \hat{\xi} + i\,\hat{\eta} = \frac{\tau - \tau_1}{\tau - \tau_2}\,.$$

(52) $$\hat{\tau} = \operatorname{Tg}\frac{\varrho}{2} \cdot e^{i\vartheta} \quad \text{für } t > 0\,; \quad |\hat{\tau}| = e^{\vartheta}, \frac{\hat{\xi}}{\hat{\eta}} = \operatorname{Sin}\varrho \quad \text{für } t < 0\,.$$

Einfache Rechnungen ergeben

(53) $$dv = \begin{cases} \operatorname{Sin}\varrho\; d\varrho\, d\vartheta & \text{für } t > 0\,, \\ \operatorname{Cos}\varrho\; d\varrho\, d\vartheta & \text{für } t < 0 \end{cases}$$

und

(54) $$\Delta_S = \begin{cases} \dfrac{1}{\operatorname{Sin}\varrho}\left\{\dfrac{\partial}{\partial\varrho}\left(\operatorname{Sin}\varrho\,\dfrac{\partial}{\partial\varrho}\right) + \dfrac{1}{\operatorname{Sin}\varrho}\,\dfrac{\partial^2}{\partial\vartheta^2}\right\} & \text{für } t > 0\,, \\ \dfrac{1}{\operatorname{Cos}\varrho}\left\{\dfrac{\partial}{\partial\varrho}\left(\operatorname{Cos}\varrho\,\dfrac{\partial}{\partial\varrho}\right) + \dfrac{1}{\operatorname{Cos}\varrho}\,\dfrac{\partial^2}{\partial\vartheta^2}\right\} & \text{für } t < 0\,. \end{cases}$$

Offenbar hat ϱ in (45) und (52) ein und dieselbe Bedeutung.

§ 2. Der Laplace-Beltramische Differentialoperator

Eine für $S[X] > 0$ erklärte Funktion $\varphi(X)$ stellt eine Funktion über $\mathfrak{P}(S)$ dar, wenn $\varphi(XV) = \varphi(X)$ für nicht-singuläres V gilt. $\Omega_1 \cong \Omega_2$ für zwei Differentialoperatoren Ω_1 und Ω_2 bedeute, daß diese auf Funktionen über $\mathfrak{P}(S)$ dieselbe Wirkung haben. In diesem Sinne ist

$$\sigma(L^2) \cong 2\sigma\left\{(E - YY')\left((E - Y'Y)\frac{\partial}{\partial Y'}\right)'\frac{\partial}{\partial Y'}\right\} - 4\sigma\left\{(E - YY')Y\frac{\partial}{\partial Y'}\right\} \tag{55}$$

mit

$$L = SX\frac{\partial}{\partial X'}S^{-1} - \left(X\frac{\partial}{\partial X'}\right)'. \tag{56}$$

Zum Beweis wird

$$dX = C\begin{pmatrix}0\\ -dY'\end{pmatrix}H + C\begin{pmatrix}E\\ -Y'\end{pmatrix}dH$$

in $d\varphi = \sigma\left(dX'\frac{\partial}{\partial X}\right)\varphi$ eingetragen. Man erhält auf diese Weise

$$\frac{\partial}{\partial Y} = -H\frac{\partial}{\partial X'}C\begin{pmatrix}0\\ E\end{pmatrix}; \qquad \frac{\partial}{\partial H} = (E, -Y)C'\frac{\partial}{\partial X}$$

und damit

$$C'\frac{\partial}{\partial X} = \begin{pmatrix}E & -Y\\ 0 & -E\end{pmatrix}\begin{pmatrix}\dfrac{\partial}{\partial H}\\ \dfrac{\partial}{\partial Y'}H'^{-1}\end{pmatrix}.$$

Wegen $\frac{\partial}{\partial H} \cong 0$ folgt

$$\frac{\partial}{\partial X'}C \cong -H^{-1}\left(\left(Y\frac{\partial}{\partial Y'}\right)', \frac{\partial}{\partial Y}\right)$$

und

$$C^{-1}X\frac{\partial}{\partial X'}C \cong \begin{pmatrix}-E\\ Y'\end{pmatrix}\left(\left(Y\frac{\partial}{\partial Y'}\right)', \frac{\partial}{\partial Y}\right),$$

mithin

$$SX\frac{\partial}{\partial X'}S^{-1} \cong -C'^{-1}\begin{pmatrix}E\\ Y'\end{pmatrix}\left(\left(Y\frac{\partial}{\partial Y'}\right)', -\frac{\partial}{\partial Y}\right)C',$$

$$\left(X\frac{\partial}{\partial X'}\right)' \cong -C'^{-1}\begin{pmatrix}Y\dfrac{\partial}{\partial Y'} & -Y\left(Y'\dfrac{\partial}{\partial Y}\right)'\\ \dfrac{\partial}{\partial Y'} & -\left(Y'\dfrac{\partial}{\partial Y}\right)'\end{pmatrix}C'.$$

Differenzbildung führt auf

$$L \cong C'^{-1}\begin{pmatrix}Y\dfrac{\partial}{\partial Y'} - \left(Y\dfrac{\partial}{\partial Y'}\right)' & \dfrac{\partial}{\partial Y} - Y\left(Y'\dfrac{\partial}{\partial Y}\right)'\\ \dfrac{\partial}{\partial Y'} - Y'\left(Y\dfrac{\partial}{\partial Y'}\right)' & Y'\dfrac{\partial}{\partial Y} - \left(Y'\dfrac{\partial}{\partial Y}\right)'\end{pmatrix}C',$$

woraus

$$\sigma(L^2) \cong \sigma\left(Y\frac{\partial}{\partial Y'} - \left(Y\frac{\partial}{\partial Y'}\right)'\right)^2 + \sigma\left(Y'\frac{\partial}{\partial Y} - \left(Y'\frac{\partial}{\partial Y}\right)'\right)^2 +$$
$$+ 2\sigma\left(\frac{\partial}{\partial Y} - Y\left(Y'\frac{\partial}{\partial Y}\right)'\right)\left(\frac{\partial}{\partial Y'} - Y'\left(Y\frac{\partial}{\partial Y'}\right)'\right)$$

erhellt. Mit Hilfe der Vertauschungsregeln

$$\left(Y\frac{\partial}{\partial Y'}\right)' Y = \left(Y' Y\frac{\partial}{\partial Y'}\right)' + Y, \left(Y'\frac{\partial}{\partial Y}\right)' Y' = \left(Y Y'\frac{\partial}{\partial Y}\right)' + Y'$$

$$\frac{\partial}{\partial Y'} Y = \left(Y'\frac{\partial}{\partial Y}\right)' + nE, \quad \frac{\partial}{\partial Y} Y' = \left(Y\frac{\partial}{\partial Y'}\right)' + (m-n)\, Y,$$

$$\sigma\left(Y\frac{\partial}{\partial Y'}\right)^2 = \sigma\left(Y\left(Y'\frac{\partial}{\partial Y}\right)'\frac{\partial}{\partial Y'}\right) + n\sigma\left(Y'\frac{\partial}{\partial Y}\right),$$

$$\sigma\left(Y'\frac{\partial}{\partial Y}\right)^2 = \sigma\left(Y'\left(Y\frac{\partial}{\partial Y'}\right)'\frac{\partial}{\partial Y}\right) + (m-n)\sigma\left(Y'\frac{\partial}{\partial Y}\right),$$

$$\sigma\left(\frac{\partial}{\partial Y} Y'\left(Y\frac{\partial}{\partial Y'}\right)'\right) = \sigma\left(Y\left(Y'\frac{\partial}{\partial Y}\right)'\frac{\partial}{\partial Y'}\right) + m\,\sigma\left(Y'\frac{\partial}{\partial Y}\right)$$

läßt sich dieser Ausdruck schließlich in (55) überführen.

Die Transformation $X \to UX\,(U \in \Omega(S))$ bewirkt $L \to U'^{-1} L U'$, läßt also $\sigma(L^2)$ invariant. An der Stelle $Y = 0$ ist der Operator $\frac{1}{2}\sigma(L^2)$ und zufolge $ds^2 = \sigma(d\,Y'\,d\,Y)$ auch der Laplace-Beltramische Operator Δ_S mit $\sigma\left(\frac{\partial}{\partial Y'}\frac{\partial}{\partial Y}\right)$ identisch. Da $\mathfrak{P}(S)$ ein homogener Raum und auch Δ_S ein invarianter Operator ist, gilt

$$\Delta_S \cong \frac{1}{2}\sigma(L^2) \tag{57}$$

allgemein.

Um die Wirkung von $\sigma(L^2)$ auf eine beliebige Funktion von $w = (S[X])^{-1}[X' S \mathfrak{g}]$ zu ermitteln, führen wir

$$T = (X'X)^{-1}[X'], \quad \Lambda = X\frac{\partial}{\partial X'} - \left(X\frac{\partial}{\partial X'}\right)' \tag{58}$$

ein und stellen zunächst fest, daß

$$X\frac{\partial}{\partial X'}\varphi(T) = 2\left(T\frac{\partial}{\partial T} - T\left(T\frac{\partial}{\partial T}\right)'\right)\varphi(T)$$

mit der für symmetrische T üblichen Definition

$$T = (t_{\alpha\beta}), \frac{\partial}{\partial T} = \left(e_{\alpha\beta}\frac{\partial}{\partial t_{\alpha\beta}}\right), e_{\alpha\beta} = \frac{1}{2}(\delta_{\alpha\beta} + 1)$$

identisch in X gilt. Mithin ist

$$\Lambda\varphi(T) = 2\left(T\frac{\partial}{\partial T} - \left(T\frac{\partial}{\partial T}\right)'\right)\varphi(T).$$

Speziell für $\varphi(T) = T[\mathfrak{p}]$ mit konstanter Spalte $\mathfrak{p}$ ergibt sich

$$\Lambda(T[\mathfrak{p}]) = 2(T\mathfrak{p}\mathfrak{p}' - \mathfrak{p}\mathfrak{p}'T).$$

Beachtet man noch, daß $\sigma(T) = n$ und daher

$$\left(T\frac{\partial}{\partial T}\right) T = \frac{n}{2}E + \frac{1}{2}T \quad \text{sowie} \quad \frac{\partial}{\partial T} T = \frac{m+1}{2}E$$

und mit konstanter Matrix $A = A^{(m)}$ auch

$$\frac{\partial}{\partial T} A T = \frac{1}{2}A' + \frac{1}{2}\sigma(A)E, \quad \left(T\frac{\partial}{\partial T}\right)' A T = \frac{1}{2}\sigma(TA)E + \frac{1}{2}A'T$$

ist, so folgt

$$\Lambda^2 (T[\mathfrak{p}]) = 4\left(T\frac{\partial}{\partial T} - \left(T\frac{\partial}{\partial T}\right)'\right)(T\mathfrak{p}\mathfrak{p}' - \mathfrak{p}\mathfrak{p}'T)$$
$$= 2(m-1)\,T\mathfrak{p}\mathfrak{p}' + 2\mathfrak{p}\mathfrak{p}'T + 2(T[\mathfrak{p}])E - 2(\mathfrak{p}'\mathfrak{p})T - 2n\mathfrak{p}\mathfrak{p}'$$

und somit

$$\sigma(\Lambda^2)(T[\mathfrak{p}]) = 4(mT[\mathfrak{p}] - n\mathfrak{p}'\mathfrak{p}). \tag{59}$$

Wegen $T^2 = T$ ist ferner

$$(\Lambda(T[\mathfrak{p}]))^2 = 4(T\mathfrak{p}\mathfrak{p}' - \mathfrak{p}\mathfrak{p}'T)^2$$
$$= 4\{(T\mathfrak{p}\mathfrak{p}' + \mathfrak{p}\mathfrak{p}'T - \mathfrak{p}\mathfrak{p}')T[\mathfrak{p}] - T\mathfrak{p}\mathfrak{p}'T(\mathfrak{p}'\mathfrak{p})\},$$

also

$$\sigma(\Lambda(T[\mathfrak{p}]))^2 = 8T[\mathfrak{p}]\,(T[\mathfrak{p}] - \mathfrak{p}'\mathfrak{p}). \tag{60}$$

Ist $S = Q'Q$, $Q = Q^{(m)}$ und wird die Substitution $X \to QX$, $\mathfrak{p} \to Q\mathfrak{g}$ ausgeführt, demzufolge auch $\Lambda \to Q'^{-1}LQ'$, $\mathfrak{p}'\mathfrak{p} \to S[\mathfrak{g}] = t$, $T[\mathfrak{p}] \to w$, so gehen (59) und (60) in

$$\sigma(L^2)w = 4(mw - nt), \quad \sigma(Lw)^2 = 8w(w-t) \tag{61}$$

über. Damit erhalten wir die gewünschte Formel

$$\begin{aligned}\sigma(L^2)\,\varphi(w) &= \sigma(L\varphi'(w)\,Lw)\\ &= \varphi''(w)\,\sigma(Lw)^2 + \varphi'(w)\,\sigma(L^2)w\\ &= 8w(w-t)\,\varphi''(w) + 4(mw - nt)\,\varphi'(w),\end{aligned} \tag{62}$$

insbesondere also

$$\begin{aligned}&\sigma(L^2)\,e^{-2\pi y w}\\ &= 16\pi y\left\{2\pi y w^2 - \left(2\pi t y + \frac{m}{2}\right)w + \frac{n}{2}t\right\}e^{-2\pi y w}.\end{aligned} \tag{63}$$

§ 3. Eine Greensche Formel

Im folgenden sei $[dX]$ das alternierende Produkt der Differentiale $dx_{\alpha\beta}$ aller Elemente $x_{\alpha\beta}$ von X in festgelegter Reihenfolge. $\omega_{\alpha\beta}$ entstehe aus $[dX]$, vom Vorzeichen abgesehen, durch Streichen von $dx_{\alpha\beta}$; das Vorzeichen von $\omega_{\alpha\beta}$ sei durch $[dX] = dx_{\alpha\beta}\,\omega_{\alpha\beta}$ festgelegt. Sind $A = A^{(m)}$ und $B = B^{(m)}$ Matrizen mit Funktionen von X als Elementen und ist Λ der durch (58) eingeführte Operator, so gilt mit $\Omega = (\omega_{\alpha\beta})$, wie in [5] gezeigt wurde,

$$\begin{aligned}&A\Lambda^2 B[dX] - (B'\Lambda^2 A')'[dX]\\ &= d\{A(X\Omega' - \Omega X')(\Lambda B) + (\Lambda A')'(X\Omega' - \Omega X')B\}.\end{aligned} \tag{64}$$

Sind $p_{\alpha\beta}$ die Elemente der Matrix $P = X'X$, so daß

$$dp_{\alpha\beta} = \sum_{\varrho}(x_{\varrho\beta}dx_{\varrho\alpha} + x_{\varrho\alpha}dx_{\varrho\beta})$$

gilt, so folgt gemäß der Definition der $\omega_{\alpha\beta}$ leicht

$$dp_{\alpha\beta}(X\Omega' - \Omega X') = 0.$$

Die Elemente der Matrix $X\Omega' - \Omega X'$ sind demnach durch das Produkt $[dP] = \prod_{\alpha \leq \beta} dp_{\alpha\beta}$ teilbar; denn die Funktionen $p_{\alpha\beta} (\alpha \leq \beta)$ sind unabhängig,

lassen sich also durch $mn - \frac{n(n+1)}{2}$ weitere Funktionen zu einem System von mn unabhängigen Funktionen von X ergänzen. Es gibt also eine Matrix M mit Differentialformen vom Grad $mn - \frac{n(n+1)}{2} - 1$ als Elementen, so daß

$$X\Omega' - \Omega X' = [dP]M$$

wird. (64) wird nun nach bekannter Regel in

$$\begin{aligned} A\Lambda^2 B[dX] &- (B'\Lambda^2 A')'[dX] \\ &= (-1)^{\frac{n(n+1)}{2}}[dP]\,d\{AM(\Lambda B) + (\Lambda A')'MB\} \end{aligned} \tag{65}$$

übergeführt. Die Substitution $X \to QX$ mit einem durch $S = Q'Q$ bestimmten quadratischen Q bewirke $M \to M^*$; im übrigen hat sie

$$[dX] \to |Q|^n[dX],\ \Lambda \to Q'^{-1}LQ',\ P \to W = S[X]$$

zur Folge. Die spezielle Wahl $A = |Q|^{-\frac{n}{2}}Q'\varphi$ und $B = |Q|^{-\frac{n}{2}}Q'^{-1}\psi$ mit willkürlichen Funktionen $\varphi = \varphi(X)$, $\psi = \psi(X)$ ergibt damit nach erfolgter Spurbildung

$$\{\varphi\sigma(L^2)\psi - \psi\sigma(L^2)\varphi\}\,[dX] = [dW]\,d\sigma\{N(\varphi L\psi - \psi L\varphi)\} \tag{66}$$

mit

$$N = (-1)^{\frac{n(n+1)}{2}}|Q|^{-n}Q'M^*Q'^{-1},$$

wobei noch von $(S^{-1}LS)' = -L$ Gebrauch gemacht wurde. Da der Kalkül der alternierenden Differentialformen gegenüber Variablensubstitutionen invariant ist und die Beschränkung auf $W =$ konstant einer solchen Substitution entspricht, so folgt nun

$$\{\varphi\sigma(L^2)\psi - \psi\sigma(L^2)\varphi\}\,\frac{[dX]}{[dW]} = d\sigma\{N(\varphi L\psi - \psi L\varphi)\}\,. \tag{67}$$

Dieser Ausdruck ist gegenüber der Substitution $X \to UX$ mit $S[U] = S, |U| = 1$ invariant, wenn dies für die Funktionen $\varphi(X)$, $\psi(X)$ zutrifft. Die Anwendung der Stokesschen Integralformel ergibt schließlich

$$\int\limits_{\substack{S[X]=W \\ X\in\mathfrak{B}_0}} \{\varphi\sigma(L^2)\psi - \psi\sigma(L^2)\varphi\}\,\frac{[dX]}{[dW]} = \int\limits_{\mathfrak{R}_0} \sigma\{N(\varphi L\psi - \psi L\varphi)\} \tag{68}$$

für einen beliebigen Bereich $\mathfrak{B}_0$ im X-Raum, wenn wir mit $\mathfrak{R}_0$ den Rand des Durchschnitts von $\mathfrak{B}_0$ mit $S[X] = W$ in bezug auf $S[X] = W$ bezeichnen, sofern nur $\mathfrak{R}_0$ hinreichend glatt ist und die Integrale existieren.

§ 4. Der Gammafaktor

Es sei $W_{\varkappa,\mu}(y)$ die Whittakersche Funktion,

$$W(y;\alpha,\beta,\varepsilon) = y^{-\frac{\alpha+\beta}{2}}\,W_{\frac{(\alpha-\beta)\varepsilon}{2},\,\frac{\alpha+\beta-1}{2}}(2y) \quad (\varepsilon = \pm 1) \tag{69}$$

und

$$\Gamma(s;\alpha,\beta)=\int_0^\infty W(y;\alpha,\beta,1)\,y^{s-1}dy\,.$$

Diese Funktion trat als „Gammafaktor" in Funktionalgleichungen Dirichletscher Reihen bereits an anderer Stelle [6] in Erscheinung. Dort wurde auch

$$\Gamma(s;\alpha,\beta)=2^{\frac{\alpha-\beta}{2}}\frac{\Gamma(s)\,\Gamma(s+1-\alpha-\beta)}{\Gamma(s+1-\alpha)}F\left(\beta,1-\alpha;s+1-\alpha;\frac{1}{2}\right) \tag{70}$$

bewiesen, wobei $F(\alpha,\beta;\gamma;x)$ die hypergeometrische Funktion bezeichnet. Die Anwendung der Identität

$$F(\alpha,\beta;\gamma;x)=(1-x)^{\gamma-\alpha-\beta}F(\gamma-\alpha,\gamma-\beta;\gamma;x)$$

ergibt

$$\begin{aligned}&\Gamma(s;\alpha,\beta)\\ &=2^{\frac{\alpha+\beta}{2}-s}\frac{\Gamma(s)\,\Gamma(s+1-\alpha-\beta)}{\Gamma(s+1-\alpha)}F\left(s,s+1-\alpha-\beta;s+1-\alpha;\frac{1}{2}\right).\end{aligned} \tag{71}$$

Zu einer Funktionalgleichung für $\Gamma(s;\alpha,\beta)$ gelangen wir auf Grund der Differentiationsformeln

$$F'(\alpha,\beta;\gamma;x)=\frac{\alpha\beta}{\gamma}F(\alpha+1,\beta+1;\gamma+1;x)\,,$$

$$F''(\alpha,\beta;\gamma;x)=\frac{\alpha(\alpha+1)\,\beta(\beta+1)}{\gamma(\gamma+1)}F(\alpha+2,\beta+2;\gamma+2;x)\,,$$

die in ersichtlicher Weise

$$\begin{aligned}&\Gamma(s+1;\alpha,\beta)\\ &=2^{\frac{\alpha+\beta}{2}-s-1}\frac{\Gamma(s)\,\Gamma(s+1-\alpha-\beta)}{\Gamma(s+1-\alpha)}F'\left(s,s+1-\alpha-\beta;s+1-\alpha;\frac{1}{2}\right),\\ &\Gamma(s+2;\alpha,\beta)\\ &=2^{\frac{\alpha+\beta}{2}-s-2}\frac{\Gamma(s)\,\Gamma(s+1-\alpha-\beta)}{\Gamma(s+1-\alpha)}F''\left(s,s+1-\alpha-\beta;s+1-\alpha;\frac{1}{2}\right)\end{aligned}$$

ergeben, wenn wir noch die Differentialgleichung der hypergeometrischen Funktion ins Treffen führen. Derzufolge ist

$$\begin{aligned}F''\left(s,s+1-\alpha-\beta;s+1-\alpha;\frac{1}{2}\right)+2(\beta-\alpha)F'\left(s,s+1-\alpha-\beta;s+1-\alpha;\frac{1}{2}\right)-\\ -4s(s+1-\alpha-\beta)F\left(s,s+1-\alpha-\beta;s+1-\alpha;\frac{1}{2}\right)=0\,,\end{aligned}$$

woraus

$$\Gamma(s+2;\alpha,\beta)+(\beta-\alpha)\Gamma(s+1;\alpha,\beta)-s(s+1-\alpha-\beta)\Gamma(s;\alpha,\beta)=0 \tag{72}$$

erhellt.

Eine Berechnung der Determinante

$$D(s;\alpha,\beta)=\begin{vmatrix}\Gamma(s;\alpha,\beta) & \Gamma(s;\beta,\alpha)\\ -\Gamma(s+1;\alpha,\beta) & \Gamma(s+1;\beta,\alpha)\end{vmatrix} \tag{73}$$

wird durch den Umstand begünstigt, daß diese Funktion der Funktionalgleichung

$$D(s+1;\alpha,\beta) = s(s+1-\alpha-\beta)\,D(s;\alpha,\beta) \tag{74}$$

genügt. Nach (72) ist in der Tat

$$D(s+1;\alpha,\beta) = \begin{vmatrix} \Gamma(s+1;\alpha,\beta) & \Gamma(s+1;\beta,\alpha) \\ -\Gamma(s+2;\alpha,\beta) & \Gamma(s+2;\beta,\alpha) \end{vmatrix}$$

$$= s(s+1-\alpha-\beta)\begin{vmatrix} \Gamma(s+1;\alpha,\beta) & \Gamma(s+1;\beta,\alpha) \\ -\Gamma(s;\alpha,\beta) & \Gamma(s;\beta,\alpha) \end{vmatrix} = s(s+1-\alpha-\beta)\,D(s;\alpha,\beta).$$

Man stellt nun fest, daß

$$H(s) = \frac{D(s;\alpha,\beta)}{\Gamma(s)\,\Gamma(s+1-\alpha-\beta)}$$

eine periodische Funktion mit der Periode 1 ist. Da $H(s)$ für hinreichend große Werte von $\sigma = \Re s$ regulär ist, ist $H(s)$ eine ganze Funktion von s. Wir zeigen, daß $\lim\limits_{\sigma\to\infty} H(s) = 2$ ist, woraus dann $H(s) = 2$ identisch in s folgt. Die Existenz des Grenzwerts ist evident auf Grund der Darstellung

$$H(s) = \frac{\Gamma(s)\,\Gamma(s+2-\alpha-\beta)}{\Gamma(s+1-\alpha)\,\Gamma(s+1-\beta)}\cdot$$

$$\cdot\begin{vmatrix} F\left(\beta,1-\alpha;s+1-\alpha;\frac{1}{2}\right) & F\left(\alpha,1-\beta;s+1-\beta;\frac{1}{2}\right) \\ -\frac{s}{s+1-\alpha}F\left(\beta,1-\alpha;s+2-\alpha;\frac{1}{2}\right) & \frac{s}{s+1-\beta}F\left(\alpha,1-\beta;s+2-\beta;\frac{1}{2}\right) \end{vmatrix},$$

die mit Hilfe von (70) gewonnen wird. Denn es ist

$$\lim_{\sigma\to\infty}\frac{\Gamma(s+a)}{\Gamma(s)}\,e^{-a\log s} = \lim_{\sigma\to\infty} F\left(\alpha,\beta;s;\frac{1}{2}\right) = 1$$

und daher $\lim\limits_{\sigma\to\infty} H(s) = 2$, wie behauptet wurde. Damit ist auch

$$D(s;\alpha,\beta) = 2\,\Gamma(s)\,\Gamma(s+1-\alpha-\beta) \tag{75}$$

bewiesen.

§ 5. Thetareihen

Um die Bildung des Integralmittelwertes $g(z;S,u)$ der Funktion $f(z;S,P)\,u(X)$ explizit auszuführen, bedürfen wir einer geeigneten Umformung der durch (1) erklärten Thetareihe, wobei zunächst wie in [8] verfahren wird. Die Anwendung von (3) ergibt die Aufspaltung

$$\begin{aligned} f(z;S,P) &= \sum_{\mathfrak{g}} e^{\pi i\bar{z}S[\mathfrak{g}]-2\pi yK[\mathfrak{g}]} \\ &= \sum_{\mathfrak{g}\,l\,\Gamma(S)} e^{\pi i\bar{z}S[\mathfrak{g}]} \sum_{\substack{U\in\Gamma(S)\\ U\,r\,\Gamma(S,\mathfrak{g})}} e^{-2\pi yK[U\mathfrak{g}]}. \end{aligned} \tag{76}$$

Hierin bedeutet $\mathfrak{g}\,l\,\Gamma(S)$ die Summation über ein volles Repräsentantensystem der Linksrestklassen $\Gamma(S)\mathfrak{g}$ ganzer Spalten und $U\,r\,\Gamma(S,\mathfrak{g})$ die Summation über ein volles Repräsentantensystem der Rechtsrestklassen $U\,\Gamma(S,\mathfrak{g})$ (in $\Gamma(S)$), wobei $\Gamma(S,\mathfrak{g})$ die Gruppe

$$\Gamma(S,\mathfrak{g}) = \{U : U\in\Gamma(S),\ U\mathfrak{g} = \mathfrak{g}\}$$

bezeichnet. Durchläuft U ein volles Repräsentantensystem der Rechtsrestklassen $U\Gamma(S,\mathfrak{g})$ in $\Gamma(S)$, so bildet die Vereinigung der Bereiche $\mathfrak{F}(S)[U]$ einen einfach überdeckten Fundamentalbereich $\mathfrak{F}(S,\mathfrak{g})$ von $\Gamma(S,\mathfrak{g})$ in $\mathfrak{P}(S)$, wenn $\mathfrak{g}=0$ oder $-E \notin \Gamma(S)$ ist. Eine doppelte Überdeckung findet im Falle $\mathfrak{g} \neq 0$, $-E \in \Gamma(S)$ statt. Mit

$$\delta_S(\mathfrak{g}) = \begin{cases} 2, & \text{falls } \mathfrak{g} \neq 0,\ -E \in \Gamma(S) \\ 1 & \text{sonst} \end{cases} \tag{77}$$

wird also

$$\begin{aligned} g(z;S,u) &= I\{f(z;S,P)\,u(X)\} \\ &= \sum_{\mathfrak{g}/\Gamma(S)} \delta_S(\mathfrak{g})\, e^{\pi i \bar{z} S[\mathfrak{g}]} \frac{1}{V(S)} \int\limits_{\mathfrak{F}(S,\mathfrak{g})} e^{-2\pi y K[\mathfrak{g}]} u(X)\, dv \end{aligned} \tag{78}$$

unter der Voraussetzung, daß $u(X)$ den Bedingungen (7) genügt. Wir verwenden im folgenden die Bezeichnungen

$$\begin{aligned} &W = S[X]\,, \quad \mathfrak{w} = X'S\mathfrak{g}\,, \quad w = W^{-1}[\mathfrak{w}] = K[\mathfrak{g}]\,, \\ &X = CX_0\,, \quad X_0 = \begin{pmatrix} 0 \\ -Y' \end{pmatrix} H\,, \quad W_0 = E - YY'\,, \quad S[\mathfrak{g}] = t\,. \end{aligned} \tag{79}$$

Um eine Integration im Y-Raum in eine solche auf $W=$ konstant im X-Raum zu verwandeln, führen wir zunächst die Umrechnung der Volumenelemente aus. Auf Grund der Differentialrelationen

$$dX = C\,dX_0\,, \quad dX_0 = \begin{pmatrix} 0 \\ -dY'H \end{pmatrix} + \begin{pmatrix} E \\ -Y' \end{pmatrix} dH'$$

$$[dX] = \|C\|^n [dX_0] = \|S\|^{-\frac{n}{2}} \|H\|^{m-n} [dY]\,[dH]\,,$$

$$\begin{aligned} dv\,\|H\|^{-n}[dH] &= |E - YY'|^{-\frac{m}{2}} \|H\|^{-n} [dY]\,[dH] \\ &= |S[X]|^{-\frac{m}{2}} \|H\|^{m-n} [dY]\,[dH] = \|S\|^{\frac{n}{2}} |W|^{-\frac{m}{2}} [dX]\,, \end{aligned}$$

also

$$\|S\|^{\frac{n}{2}} |W|^{\frac{n+1-m}{2}} \frac{[dX]}{[dW]} = dv\,|W_0|^{\frac{n}{2}} |W|^{\frac{1}{2}} \frac{[dH]}{[dW]} \tag{80}$$

ist. Ferner gilt nach Hilfssatz 3 in [8] S. 31

$$\int\limits_{W_0[H]=W} \frac{[dH]}{[dW]} = \varrho_n |W_0|^{-\frac{n}{2}} |W|^{-\frac{1}{2}} \tag{81}$$

mit

$$\varrho_n = \prod_{k=1}^{n} \frac{\pi^{\frac{k}{2}}}{\Gamma\left(\frac{k}{2}\right)}\,.$$

Sei nun $\mathfrak{B}$ ein beliebiger Bereich im Y-Raum, $\mathfrak{B}_0$ die Menge aller X, denen ein Y in $\mathfrak{B}$ entspricht und $\varphi(X)$ eine willkürliche Funktion über $\mathfrak{B}$, so daß also

$\varphi(XV) = \varphi(X)$ für $|V| \neq 0$ gilt. Dann ist

$$\int\limits_{Y\in\mathfrak{B}} \varphi(X)\,dv = \frac{1}{\varrho_n}\,\|S\|^{\frac{n}{2}}\,|W|^{\frac{n+1-m}{2}} \int\limits_{\substack{S[X]=W\\ X\in\mathfrak{B}_0}} \varphi(X)\,\frac{[dX]}{[dW]}\,; \tag{82}$$

denn die rechte Seite dieser Gleichung ist mit Hilfe von (80) und (81) in

$$\frac{1}{\varrho_n}\,|W|^{\frac{1}{2}} \int\limits_{Y\in\mathfrak{B}} \varphi(X)\,|W_0|^{\frac{n}{2}} \int\limits_{W_0[H]=W} \frac{[dH]}{[dW]}\cdot dv = \int\limits_{Y\in\mathfrak{B}} \varphi(X)\,dv$$

überzuführen, q. e. d.

Wir wenden (82) auf das Integral in (78) an, setzen also $\mathfrak{B} = \mathfrak{F}(S, \mathfrak{g})$ und $\varphi(X) = e^{-2\pi y w}u(X)$. An Stelle von $\mathfrak{B}_0$ tritt dann ein Fundamentalbereich $\mathfrak{F}_0(S, \mathfrak{g})$ in $S[X] > 0$ bezüglich der Gruppe der Substitutionen $X \to UX$ $(U \in \Gamma(S, \mathfrak{g}))$. Damit ergibt sich

$$g(z;\,S,u) = \sum_{\mathfrak{g}/\Gamma(S)} \delta_S(\mathfrak{g})\, e^{\pi i \bar{z} t}\,\frac{\|S\|^{\frac{n}{2}}}{\varrho_n V(S)}\,\chi(y;\,S,\mathfrak{g},u)\,. \tag{83}$$

Hier ist zur Abkürzung

$$\chi(y) = \chi(y;\,S,\mathfrak{g},u) = |W|^{\frac{n+1-m}{2}} \int\limits_{\substack{S[X]=W\\ X\in\mathfrak{F}_0(S,\mathfrak{g})}} e^{-2\pi y w}u(X)\,\frac{[dX]}{[dW]} \tag{84}$$

gesetzt. Zu einer Differentialgleichung für $\chi(y)$ gelangt man durch Bildung des Ausdrucks

$$\begin{aligned} &\Big\{y^2\chi''(y) + y\Big(2\pi t y + \frac{m}{2}\Big)\chi'(y) + \pi t y n\,\chi(y)\Big\}\,|W|^{\frac{m-n-1}{2}} \\ &= \int\limits_{\substack{S[X]=W\\ X\in\mathfrak{F}_0(S,\mathfrak{g})}} \Big\{4\pi^2y^2w^2 - 2\pi y w\Big(2\pi t y + \frac{m}{2}\Big) + \pi t y n\Big\}\,e^{-2\pi y w}u(X)\,\frac{[dX]}{[dW]} \\ &= \frac{1}{8} \int\limits_{\substack{S[X]=W\\ X\in\mathfrak{F}_0(S,\mathfrak{g})}} u(X)\,\sigma(L^2)\,e^{-2\pi y w}\,\frac{[dX]}{[dW]}\,, \end{aligned} \tag{85}$$

wobei von (63) Gebrauch gemacht wurde. Die Funktionen $u(X)$, $e^{-2\pi y w}$ sind gegenüber den Substitutionen $X \to UX\,(U \in \Gamma(S, \mathfrak{g}))$ invariant. Es ist zweckmäßig, hier vorauszusetzen, daß $\Gamma(S)$, damit auch $\Gamma(S, \mathfrak{g})$ nur solche Einheiten U enthält, deren Determinante $|U| = 1$ ist. Die Greensche Formel (68) erlaubt nun, (85) wegen (57) in

$$\frac{1}{8} \int\limits_{\substack{S[X]=W\\ X\in\mathfrak{F}_0(S,\mathfrak{g})}} e^{-2\pi y w}\sigma(L^2)\,u(X\,\frac{[dX]}{[dW]} = -\frac{\lambda}{4}\,\chi(y)\,|W|^{\frac{m-n-1}{2}}$$

überzuführen; denn das Randintegral verschwindet, da sich der Rand des Integrationsbereiches aus paarweise bezüglich $\Gamma(S, \mathfrak{g})$ äquivalenten Stücken mit entgegengesetzter Orientierung zusammensetzt und der Integrand des Randintegrals gegenüber $\Gamma(S, \mathfrak{g})$ invariant ist. Damit ist die Differentialgleichung

$$y^2\chi''(y) + y\Big(2\pi t y + \frac{m}{2}\Big)\chi'(y) + \Big(\pi t y n + \frac{\lambda}{4}\Big)\chi(y) = 0 \tag{86}$$

bewiesen. Sie besagt im Falle $\mathfrak{g} = 0$, daß

$$(87) \qquad \lambda I\{u\} = 0$$

ist. Fortan sei $\mathfrak{g} \neq 0$.

Im Falle $t = 0$ ist, wie man leicht bestätigt, die allgemeine Lösung von (86) in der Form

$$(88) \qquad \chi(y) = a_0 y^{\frac{\nu}{2}} W_0\left(y, \frac{m}{2} + \nu\right) + b_0 y^{\frac{\nu}{2}}$$

mit gewissen Konstanten a_0, b_0 darstellbar, wobei ν die Lösung von

$$(89) \qquad \nu(\nu + m - 2) + \lambda = 0\,, \quad \Re e\, \nu \geq \frac{2-m}{2}\,, \quad \Im m\, \nu \geq 0$$

und

$$(90) \qquad W_0(y, \gamma) = \frac{y^{1-\gamma} - 1}{1-\gamma} = \sum_{\alpha=1}^{\infty} \frac{(\log y)^\alpha}{\alpha!} (1-\gamma)^{\alpha-1}$$

gesetzt ist.

Im Falle $t \neq 0$ ergibt der Ansatz

$$(91) \qquad \chi(y) = a(\pi|t|\,y)^{\frac{\nu}{2}} e^{-\pi t y} W\left(\pi|t|\,y; \frac{n+\nu}{2}, \frac{m+n-\nu}{2}, \operatorname{sgn} t\right)$$

mit konstantem a für die Funktion

$$W(y) = W(y; \alpha, \beta, \varepsilon) \qquad (\varepsilon^2 = 1)$$

zufolge (86) die Differentialgleichung

$$(92) \qquad y W''(y) + (\alpha + \beta) W'(y) + ((\alpha - \beta)\varepsilon - y) W(y) = 0\,.$$

Wir werden damit wie in [6] auf die schon in § 4 genannte Funktion

$$(93) \qquad W(y; \alpha, \beta, \varepsilon) = y^{-\frac{\alpha+\beta}{2}} W_{\frac{(\alpha-\beta)\varepsilon}{2}, \frac{\alpha+\beta-1}{2}}(2y)$$

geführt. Eine andere, von (93) linear unabhängige Lösung von (92) kann wegen ihres Verhaltens für $y \to \infty$ in der Darstellung von $\chi(y)$ nicht auftreten.

Nachdem die Integrale $\chi(y)$ explizit berechnet sind, können wir nun daran gehen, für die elliptischen und hyperbolischen $\mathfrak{g}$ (d. h. also $t \neq 0$) Gewichtsfunktionen $\mu(S, \mathfrak{g}, u)$ in vernünftiger Weise so einzuführen, daß $\mu(S, \mathfrak{g}, 1)$ mit dem Siegelschen Ausdruck

$$(94) \qquad \mu(S, \mathfrak{g}) = \varrho_{m-n-1} \|S\|^{\frac{n-m}{2}} \{(w - t)\,|W|\}^{1-\frac{m-n}{2}} \int\limits_{\substack{S[X] = W \\ X' S \mathfrak{g} = \mathfrak{w} \\ X \in \mathfrak{F}_0(S, \mathfrak{g})}} \frac{[dX]}{[d\mathfrak{w}]\,[dW]}$$

übereinstimmt. Die Invarianzbetrachtungen in [8] zeigen, daß $\mu(S, \mathfrak{g})$ von $W, \mathfrak{w}$, also auch von $W^{-1}[\mathfrak{w}] = w$ nicht abhängt. Eine formale Umformung ergibt zunächst

$$(95) \qquad \chi(y) = \int\limits_{w_0}^{\infty} e^{-2\pi y w} (w - t)^{\frac{m-n}{2}-1} w^{\frac{n}{2}-1} \varphi(w)\, dw$$

mit $w_0 = \mathrm{Max}(0, t)$ und

(96) $$\varphi(w) = \varphi(w, S, \mathfrak{g}, u) = \int\limits_{\substack{S[X] = W \\ W^{-1}[X' S \mathfrak{g}] = w \\ X \in \mathfrak{F}_0(S, \mathfrak{g})}} u(X)\, \omega\,,$$

wobei ω das in gewissem Sinne invariante Volumenelement

(97) $$\omega = (w-t)^{1-\frac{m-n}{2}}\, w^{1-\frac{n}{2}}\, |W|^{\frac{n+1-m}{2}}\, \frac{[dX]}{dw\,[dW]}$$

bezeichnet. Speziell für $u = 1$ ergibt die Anwendung von Hilfssatz 3 in [8], S. 31

$$\begin{aligned}
&\varphi(w, S, \mathfrak{g}, 1) \\
&\quad = \int\limits_{W^{-1}[\mathfrak{w}] = w} \left\{ \int\limits_{\substack{S[X] = W \\ X' S \mathfrak{g} = \mathfrak{w} \\ X \in \mathfrak{F}_0(S, \mathfrak{g})}} \{(w-t)|W|\}^{1-\frac{m-n}{2}} \frac{[dX]}{[d\mathfrak{w}]\,[dW]} \right\} |W|^{-\frac{1}{2}}\, w^{\frac{n}{2}-1}\, \frac{[d\mathfrak{w}]}{dw} \\
&\quad = \frac{\mu(S, \mathfrak{g})}{\varrho_{m-n-1}} \|S\|^{\frac{m-n}{2}} \int\limits_{W^{-1}[\mathfrak{w}] = w} |W|^{-\frac{1}{2}}\, w^{\frac{n}{2}-1}\, \frac{[d\mathfrak{w}]}{dw} = \frac{\mu(S, \mathfrak{g})\, \varrho_n}{\varrho_{m-n-1}\, \varrho_{n-1}} \|S\|^{\frac{m-n}{2}}.
\end{aligned}$$

Damit ist die Definition

(98) $$\mu(S, \mathfrak{g}, u) = \frac{\varrho_{m-n-1}\, \varrho_{n-1}}{\varrho_n} \|S\|^{\frac{n-m}{2}}\, \varphi(w_0, S, \mathfrak{g}, u) \qquad (t \neq 0)$$

nahegelegt. Es bleibt zu zeigen, daß $\varphi(w)$ an der Stelle w_0 vernünftig, d. h. stetig oder zumindest stetig erklärbar ist. $\varphi(w)$ kann nun aber explizit angegeben werden. Zufolge

$$\chi(y) e^{2\pi y w_0} = \int_0^\infty e^{-2\pi y w} (w + w_0 - t)^{\frac{m-n}{2}-1} (w + w_0)^{\frac{n}{2}-1}\, \varphi(w + w_0)\, dw$$

und (91) ist

$$a \left(\frac{|t| y}{2}\right)^{\frac{\nu}{2}} e^{\frac{|t| y}{2}}\, W\left(\frac{|t| y}{2};\, \frac{n+\nu}{2},\, \frac{m-n+\nu}{2},\, \mathrm{sgn}\, t\right)$$

die Laplace-Transformierte der Funktion

(99) $$(w + w_0 - t)^{\frac{m-n}{2}-1} (w + w_0)^{\frac{n}{2}-1}\, \varphi(w + w_0)\,,$$

diese also durch erstere eindeutig bestimmt. Mit $\varepsilon = \mathrm{sgn}\, t$ gilt nun nach (93) und [1], S. 294 (9):

(100) $$\begin{aligned}
\left(\frac{|t| y}{2}\right)^{\frac{\nu}{2}} e^{\frac{|t| y}{2}}\, W\left(\frac{|t| y}{2};\, \frac{n+\nu}{2},\, \frac{m-n+\nu}{2},\, \mathrm{sgn}\, t\right)
&= \left(\frac{|t| y}{2}\right)^{-\frac{m}{4}} e^{\frac{|t| y}{2}}\, W_{\frac{(2n-m)\varepsilon}{4},\, \frac{m+2\nu-2}{4}}(|t| y) \\
&= \int_0^\infty e^{-\nu w} f(w)\, dw
\end{aligned}$$

mit

$$f(w) = 2^{\frac{m}{4}}\, |t|^{\frac{(2n-m)\varepsilon - m}{4}}\, w^{\frac{m + (m-2n)\varepsilon}{4} - 1} \left(\Gamma\left(\frac{m + (m-2n)\varepsilon}{4}\right)\right)^{-1}$$
$$F\left(\frac{(m-2n)\varepsilon + m + 2\nu}{4},\, \frac{(m-2n)\varepsilon - m - 2\nu}{4} + 1;\, \frac{m + (m-2n)\varepsilon}{4};\, -\frac{w}{t}\right).$$

Diese Funktion ist also bis auf den konstanten Faktor a mit (99) identisch. Damit erhält man für $t > 0$:

$$(w+t)^{\frac{n}{2}-1}\ \varphi(w+t)$$
$$= \frac{a 2^{\frac{m}{4}}}{\Gamma\left(\frac{m-n}{2}\right)} t^{-\frac{m-n}{2}} F\left(\frac{m-n+\nu}{2}, 1-\frac{n+\nu}{2}; \frac{m-n}{2}; -\frac{w}{t}\right)$$
$$= \frac{a 2^{\frac{m}{4}}}{\Gamma\left(\frac{m-n}{2}\right)} t^{-\frac{m-n}{2}} \left(1+\frac{w}{t}\right)^{\frac{n}{2}-1} F\left(-\frac{\nu}{2}, \frac{m+\nu}{2}-1; \frac{m-n}{2}; -\frac{w}{t}\right),$$

also

$$\varphi(w) = \frac{a 2^{\frac{m}{4}}}{\Gamma\left(\frac{m-n}{2}\right)} t^{1-\frac{m}{2}} F\left(-\frac{\nu}{2}, \frac{m+\nu}{2}-1; \frac{m-n}{2}; 1-\frac{w}{t}\right) \quad (t>0) \tag{101}$$

für $t < 0$ hingegen:

$$(w-t)^{\frac{m-n}{2}-1}\ \varphi(w)$$
$$= \frac{a 2^{\frac{m}{4}}}{\Gamma\left(\frac{n}{2}\right)} |t|^{-\frac{n}{2}} F\left(\frac{n+\nu}{2}, 1-\frac{m-n+\nu}{2}; \frac{n}{2}; \frac{w}{t}\right)$$
$$= \frac{a 2^{\frac{m}{4}}}{\Gamma\left(\frac{n}{2}\right)} |t|^{-\frac{n}{2}} \left(1-\frac{w}{t}\right)^{\frac{m-n}{2}-1} F\left(-\frac{\nu}{2}, \frac{m+\nu}{2}-1; \frac{n}{2}; \frac{w}{t}\right),$$

also

$$\varphi(w) = \frac{a 2^{\frac{m}{4}}}{\Gamma\left(\frac{n}{2}\right)} |t|^{1-\frac{m}{2}} F\left(-\frac{\nu}{2}, \frac{m+\nu}{2}-1; \frac{n}{2}; \frac{w}{t}\right) \qquad (t<0). \tag{102}$$

Allgemein ist demnach

$$a = 2^{-\frac{m}{4}} \Gamma\left(\frac{m+(m-2n)\varepsilon}{4}\right) |t|^{\frac{m}{2}-1} \varphi(w_0) \tag{103}$$

oder, wenn wir die Gewichtsfunktion einführen,

$$a = 2^{-\frac{m}{4}} |t|^{\frac{m}{2}-1} \frac{\mu(S,\mathfrak{g},u)\,\varrho_n}{\varrho_{m-n-1}\varrho_{n-1}} \|S\|^{\frac{m-n}{2}} \Gamma\left(\frac{m+(m-2n)\varepsilon}{4}\right). \tag{104}$$

Schließlich führen wir noch in Analogie zu [8], (13) das Maß

$$\mu(S) = \frac{1}{j} \varrho_n \varrho_{m-n} \|S\|^{-\frac{m+1}{2}} V(S) \tag{105}$$

mit

$$j = \begin{cases} 1 & \text{im Falle } -E \notin \Gamma(S), \\ 2 & \text{im Falle } -E \in \Gamma(S) \end{cases} \tag{106}$$

ein, so daß nach (77) im Falle $\mathfrak{g} \neq 0$ auch $j = \delta_S(\mathfrak{g})$ gilt.

Auf eine Definition der Gewichtsfunktionen $\mu(S, \mathfrak{g}, u)$ für die parabolischen $\mathfrak{g}$ soll hier verzichtet werden. Wird zur Abkürzung noch

$$M(S, t, u) = \sum_{\substack{\mathfrak{g} \mid \Gamma(S) \\ S[\mathfrak{g}] = t}} \mu(S, \mathfrak{g}, u) \quad (t \neq 0) \tag{107}$$

und

$$\alpha_t(S, u) = \frac{2^{-\frac{m}{4}} \pi^{\frac{m}{2}} \|S\|^{-\frac{1}{2}} M(S, t, u)}{\Gamma\left(\frac{m + (2n - m)\varepsilon}{4}\right) \mu(S)} \quad (t \neq 0) \tag{108}$$

gesetzt, so kann auf Grund der Formeln (83), (88), (91), (104), (105), (107), (108) mit gewissen Konstanten $\alpha_0(S, u)$, $\beta_0(S, u)$ für

$$h(z; S, u) = (\pi y)^{-\frac{\nu}{2}} g(z; S, u) \tag{109}$$

die folgende Entwicklung notiert werden:

$$\begin{aligned} h(z; S, u) &= \alpha_0(S, u)\, W_0\left(y, \frac{m}{2} + \nu\right) + \beta_0(S, u) + \\ &+ \sum_{t \neq 0} \alpha_t(S, u) |t|^{\frac{m+\nu}{2} - 1} W\left(\pi |t| y; \frac{n + \nu}{2}, \frac{m - n + \nu}{2}, \operatorname{sgn} t\right) e^{\pi i t x}. \end{aligned} \tag{110}$$

Ist $S[\mathfrak{x}]$ keine Nullform, also notwendig $m < 5$, so gilt diese Darstellung mit $\alpha_0(S, u) = 0$, $\beta_0(S, u) = I\{u\}$; denn wenn $I\{u\} \neq 0$, dann ist nach (87) $\lambda = 0$, also auch $\nu = 0$, so daß stets $I\{u\} = I\{u\} (\pi y)^{-\frac{\nu}{2}}$ ist.

Der Beweis der Funktionalgleichung

$$f\left(-\frac{1}{z}; S, P\right) = \|S\|^{-\frac{1}{2}} (-iz)^{\frac{n}{2}} (i\bar{z})^{\frac{m-n}{2}} f(z; S^{-1}, P^{-1}) \tag{111}$$

ist nach den Angaben in [8] mühelos zu erbringen, so daß hier nicht darauf eingegangen zu werden braucht. Da zur Darstellung von $\mathfrak{P}(S^{-1})$ die Matrix $\hat{X} = SX = SC\begin{pmatrix} E \\ -Y' \end{pmatrix} H$ in gleicher Weise dienen kann wie $X = C\begin{pmatrix} E \\ -Y' \end{pmatrix} H$ für $\mathfrak{P}(S)$, so erhält man aus (111) durch Bildung der Integralmittelwerte mit $\tilde{u}(\hat{X}) = u(X)$ die Transformationsformel

$$g\left(-\frac{1}{z}; S, u\right) = \|S\|^{-\frac{1}{2}} (-iz)^{\frac{n}{2}} (i\bar{z})^{\frac{m-n}{2}} g(z; S^{-1}, \tilde{u}). \tag{112}$$

Man sieht leicht ein, daß u und $\tilde{u}$ zu demselben Eigenwert λ gehören. Zusammen mit

$$\left(\frac{\pi y}{x^2 + y^2}\right)^{-\frac{\nu}{2}} = (-iz)^{\frac{\nu}{2}} (i\bar{z})^{\frac{\nu}{2}} (\pi y)^{-\frac{\nu}{2}}$$

folgt damit

$$h\left(-\frac{1}{z}; S, u\right) = \|S\|^{-\frac{1}{2}} (-iz)^{\frac{n+\nu}{2}} (i\bar{z})^{\frac{m-n+\nu}{2}} h(z; S^{-1}, \tilde{u}). \tag{113}$$

Auf Grund der Entwicklung (110) ist nach [6] bereits evident, daß $h(z, S, u)$ von dem Operator

$$y^2\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) + \frac{m - 2n}{2} i y \frac{\partial}{\partial x} + \left(\frac{m}{2} + \nu\right) y \frac{\partial}{\partial y} \tag{114}$$

annulliert wird. Damit ist auch bereits nahegelegt, die in [6] entwickelte Theorie der Dirichletreihen mit Funktionalgleichungen auf $h(z; S, u)$ anzuwenden.

Die Beziehung, die zwischen $\mathfrak{g}$ und $u(X)$ in der Gewichtsfunktion $\mu(S, \mathfrak{g}, u)$ zum Ausdruck kommt, kann nur im Falle $n = 1$ relativ einfach beschrieben werden, weil $\mathfrak{g}$ und X dann gleichartige oder duale Gebilde der projektiven Geometrie bezeichnen. Bis auf einen konstanten Faktor stimmt $\mu(S, \mathfrak{g}, u)$ im Falle $t > 0$ mit $u(\mathfrak{g})$ überein und im Falle $t < 0$ mit dem Integralmittelwert von $u(\mathfrak{x})$ über einem Fundamentalbereich auf der Polaren zu $\mathfrak{g}$ bezüglich der Gruppe $\Gamma(S, \mathfrak{g})$, sofern $n = 1$ ist. Für den ternären Fall $(m = 3)$ soll dies nun unabhängig von den bisherigen Untersuchungen näher ausgeführt werden, wobei wir uns der speziellen, in § 1 angegebenen Parameterdarstellungen des Majorantenraumes bedienen.

§ 6. Der ternäre Fall

Im folgenden wird angenommen, daß $S[\mathfrak{x}]$ eine ternäre Form der Signatur $(1, 2)$ ist, welche 0 rational nicht darstellt, womit das Auftreten parabolischer Punkte $\mathfrak{g}$ ausgeschlossen ist. $u(\mathfrak{x})$ ist nun eine eindeutige Funktion der durch (47) eingeführten komplexen Variablen $\tau = \xi + i\eta$. An Stelle von $u(\mathfrak{x})$ werden wir daher meist $u(\tau)$ schreiben. $\Gamma(S)$ wählen wir gemäß den Angaben von § 1, so daß stets $\delta_S(\mathfrak{g}) = 1$ ist. Der Einheitengruppe $\Gamma(S)$ entspricht eine Gruppe $\Gamma^*(S)$ von Matrizen $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ mit reellen a, b, c, d und $ad - bc = 1$. $\Gamma^*(S, \mathfrak{g})$ sei die Untergruppe von $\Gamma^*(S)$, welche der Einheitengruppe $\Gamma(S, \mathfrak{g})$ entspricht. Wir setzen voraus, daß $-E$ in $\Gamma^*(S)$ und $\Gamma^*(S, \mathfrak{g})$ liegt. $\Gamma^*(S)$ ist, worauf dem Sinn nach bereits in [2] hingewiesen wurde, eine Grenzkreisgruppe erster Art im Peterssonschen Sinne ohne parabolische Substitutionen, so daß es einen kompakten Fundamentalbereich $\mathfrak{F}^*(S)$ bezüglich $\Gamma^*(S)$ in der Halbebene $\eta > 0$ gibt. $\mathfrak{F}^*(S, \mathfrak{g})$ bezeichne einen Fundamentalbereich für die Gruppe $\Gamma^*(S, \mathfrak{g})$.

Ein Vergleich von (78) mit (83) zeigt, daß im vorliegenden Fall

$$\chi(y) = \chi(y; S, \mathfrak{g}, u) = \|S\|^{-\frac{1}{2}} \iint\limits_{\mathfrak{F}^*(S, \mathfrak{g})} e^{-2\pi y w} u(\tau)\, dv \tag{115}$$

ist. Die Berechnung dieses Integrals wird für elliptische und hyperbolische $\mathfrak{g}$ getrennt durchgeführt. Wir operieren in jedem Fall mit den durch (51), (52) eingeführten Normalkoordinaten.

1. Es sei $t > 0$. Dann ist $\Gamma(S, \mathfrak{g})$ eine endliche zyklische Gruppe. Ihre Ordnung — wir nennen sie $E(S, \mathfrak{g})$ — stimmt mit der halben Ordnung von $\Gamma^*(S, \mathfrak{g})$ überein. Ein erzeugendes Element von $\Gamma(S, \mathfrak{g})$ bewirkt in der hyperbolischen Ebene eine Drehung um den Winkel $2\pi/E(S, \mathfrak{g})$. Mithin ist

$$\chi(y) = \frac{\|S\|^{-\frac{1}{2}}}{E(S, \mathfrak{g})} \int\limits_0^\infty \int\limits_0^{2\pi} e^{-2\pi t y \operatorname{Cos}^2 \varrho}\, \omega(\varrho, \vartheta) \operatorname{Sin} \varrho \, d\varrho \, d\vartheta \,,$$

wenn $u(\tau) = \omega(\varrho, \vartheta)$ gesetzt wird. Wird $\omega(\varrho, \vartheta)$ in eine Fourierreihe entwickelt:

$$\omega(\varrho, \vartheta) = \sum_{n=-\infty}^{\infty} \omega_n(\mathrm{Cos}\,\varrho)\, e^{in\vartheta},$$

so ist die Integration über ϑ auszuführen. Man erhält

$$\chi(y) = \frac{2\pi \|S\|^{-\frac{1}{2}}}{E(S,\mathfrak{g})} \int_1^{\infty} e^{-2\pi t y p^2}\, \omega_0(p)\, dp.$$

Mit $\omega(\varrho, \vartheta)$ genügt auch $\omega_0(\mathrm{Cos}\,\varrho)$ der Wellengleichung, so daß zufolge (54)

$$(1-p^2)\,\omega_0''(p) - 2p\,\omega_0'(p) + \nu(\nu+1)\,\omega_0(p) = 0$$

ist. Die Substitution $p^2 = 1 - q$ führt auf die hypergeometrische Differentialgleichung

$$q(q-1)\frac{d^2\omega_0}{dq^2} + \left(\frac{3}{2}q - 1\right)\frac{d\omega_0}{dq} - \frac{\nu(\nu+1)}{4}\,\omega_0 = 0$$

mit dem Hauptsystem

$$F\left(-\frac{\nu}{2}, \frac{\nu+1}{2}; 1; q\right), \quad F\left(-\frac{\nu}{2}, \frac{\nu+1}{2}; 1; q\right)\log q.$$

Die zweite Lösung scheidet wegen ihrer Singularität in $q = 0$ aus. Wie man leicht sieht, ist $\omega_0(1) = u(\mathfrak{g})$ und folglich

$$\omega_0(p) = u(\mathfrak{g})\, F\left(-\frac{\nu}{2}, \frac{\nu+1}{2}; 1; 1-p^2\right),$$

mithin

$$\chi(y) = \frac{2\pi\|S\|^{-\frac{1}{2}}}{E(S,\mathfrak{g})}\, u(\mathfrak{g}) \int_1^{\infty} e^{-2\pi t y p^2}\, F\left(-\frac{\nu}{2}, \frac{\nu+1}{2}; 1; 1-p^2\right) dp$$

$$= \frac{\pi\|S\|^{-\frac{1}{2}}}{E(S,\mathfrak{g})}\, u(\mathfrak{g})\, e^{-2\pi t y} \int_0^{\infty} e^{-2\pi t y q}\, F\left(-\frac{\nu}{2}, \frac{\nu+1}{2}; 1; -q\right) \frac{dq}{\sqrt{1+q}}$$

$$= \frac{\pi\|S\|^{-\frac{1}{2}}}{E(S,\mathfrak{g})}\, u(\mathfrak{g})\, e^{-2\pi t y} \int_0^{\infty} e^{-2\pi t y q}\, F\left(1+\frac{\nu}{2}, \frac{1-\nu}{2}; 1; -q\right) dq$$

$$= \frac{2^{-\frac{3}{2}}\pi\|S\|^{-\frac{1}{2}}}{E(S,\mathfrak{g})}\, u(\mathfrak{g})\, (\pi t y)^{\frac{\nu}{2}}\, e^{-\pi t y}\, W\left(\pi t y; \frac{1+\nu}{2}, \frac{2+\nu}{2}, 1\right),$$

wobei von (100) Gebrauch gemacht wurde. Der Vergleich mit (91) und (104) liefert die Formel

$$\mu(S, \mathfrak{g}, u) = \frac{\pi\|S\|^{-\frac{3}{2}}\, t^{-\frac{1}{2}}}{E(S,\mathfrak{g})}\, u(\mathfrak{g}) \qquad (t > 0). \tag{116}$$

2. Es sei $t < 0$. Nun ist $\Gamma(S, \mathfrak{g})$ eine zyklische Gruppe von hyperbolischen Substitutionen. Hat $U \in \Gamma(S, \mathfrak{g})$ in Normalkoordinaten die Wirkung $\hat{\tau} \to k\hat{\tau}$,

so nennen wir $|\log k|$ mit H. Huber [3] die Verschiebungslänge von U. Die Verschiebungslänge einer erzeugenden Transformation in $\Gamma(S, \mathfrak{g})$ werde mit $l = l(S, \mathfrak{g})$ bezeichnet. Als Fundamentalbereich $\mathfrak{F}^*(S, \mathfrak{g})$ kann der Bereich $0 \leqq \vartheta < l$ genommen werden. Wir stellen $u(\tau)$ als Funktion von ϱ, ϑ dar: $u(\tau) = \omega(\varrho, \vartheta)$ und erhalten so

$$\chi(y) = \|S\|^{-\frac{1}{2}} \int_{-\infty}^{\infty} \int_{0}^{l} e^{2\pi t y \operatorname{Sin}^2 \varrho}\, \omega(\varrho, \vartheta) \operatorname{Cos} \varrho \, d\varrho \, d\vartheta \, .$$

$\omega(\varrho, \vartheta)$ gestattet offenbar eine Fourierentwicklung der Art

$$\omega(\varrho, \vartheta) = \sum_{n=-\infty}^{\infty} \omega_n(\operatorname{Sin} \varrho)\, e^{\frac{2\pi i \vartheta}{l} n} .$$

Die Integration über ϑ kann somit ausgeführt werden und ergibt

$$\chi(y) = \|S\|^{-\frac{1}{2}} l(S, \mathfrak{g}) \int_{-\infty}^{\infty} e^{2\pi t y p^2} \omega_0(p)\, dp \, .$$

Die Wellengleichung nimmt für $\omega_0(p)$ gemäß (54) die Gestalt

$$(p^2 + 1)\, \omega_0''(p) + 2p\, \omega_0'(p) - \nu(\nu + 1)\, \omega_0(p) = 0$$

an. Die Substitution $q = -p^2$ führt auf die hypergeometrische Differentialgleichung

$$q(q-1) \frac{d^2 \omega_0}{dq^2} + \left(\frac{3}{2} q - \frac{1}{2}\right) \frac{d\omega_0}{dq} - \frac{\nu(\nu+1)}{4}\, \omega_0 = 0$$

mit dem Hauptsystem

$$F\left(-\frac{\nu}{2}, \frac{\nu+1}{2}; \frac{1}{2}; q\right), \quad q^{\frac{1}{2}} F\left(1 + \frac{\nu}{2}, \frac{1-\nu}{2}; \frac{3}{2}; q\right) .$$

Da bei der Berechnung von $\chi(y)$ nur der in p gerade Bestandteil von $\omega_0(p)$ einen Beitrag liefern kann, so folgt

$$\begin{aligned}
\chi(y) &= \|S\|^{-\frac{1}{2}} l(S, \mathfrak{g})\, \omega_0(0) \int_{-\infty}^{\infty} e^{2\pi t y p^2} F\left(-\frac{\nu}{2}, \frac{\nu+1}{2}; \frac{1}{2}; -p^2\right) dp \\
&= \|S\|^{-\frac{1}{2}} l(S, \mathfrak{g})\, \omega_0(0) \int_{0}^{\infty} e^{2\pi t y q} F\left(-\frac{\nu}{2}, \frac{\nu+1}{2}, \frac{1}{2}; -q\right) \frac{dq}{\sqrt{q}} \\
&= 2^{-\frac{3}{4}} \sqrt{\pi}\, \|S\|^{-\frac{1}{2}} l(S, \mathfrak{g})\, \omega_0(0)\, (\pi |t| y)^{\frac{\nu}{2}} e^{-\pi t y} W\left(\pi |t| y; \frac{1+\nu}{2}, \frac{2+\nu}{2}, -1\right),
\end{aligned}$$

ebenfalls wieder mit Hilfe von (100). $\omega_0(0)$ kann offenbar als Integralmittelwert von $u(\mathfrak{x})$ längs der hyperbolischen Geraden $\xi = 0$, die wir schon als Polare zu $\mathfrak{g}$ bezüglich des Kegelschnitts $S[\mathfrak{x}] = 0$ erkannt haben, angesprochen werden. Wir dürfen diesen Mittelwert mit $u(\mathfrak{g})$ bezeichnen, weil $u(\mathfrak{x})$ für $S[\mathfrak{x}] < 0$ noch nicht definiert ist. Der Vergleich mit den allgemeinen Formeln (91) und (104) ergibt schließlich

$$(117) \qquad \mu(S, \mathfrak{g}, u) = \|S\|^{-\frac{1}{2}} |t|^{-\frac{1}{2}} l(S, \mathfrak{g})\, u(\mathfrak{g}) \qquad (t < 0) \, .$$

§ 7. Dirichletreihen

Jeder Funktion $h(z; S, u)$ läßt sich nach dem in [6] entwickelten Verfahren ein Paar von meromorphen Funktionen zuordnen, die in Dirichletsche Reihen entwickelbar sind und Funktionalgleichungen vom Riemannschen Typus genügen. Zur Durchführung bedürfen wir der folgenden Reihen:

$$F(y; S, u) = \sum_{t \neq 0} \alpha_t(S, u)|t|^{\frac{m+\nu}{2}-1} W\left(\pi|t|y; \frac{n+\nu}{2}, \frac{m-n+\nu}{2}, \operatorname{sgn} t\right),$$

$$G(y; S, u) = \sum_{t \neq 0} \operatorname{sgn} t\, \alpha_t(S, u)|t|^{\frac{m+\nu}{2}} W\left(\pi|t|y; \frac{n+\nu}{2}, \frac{m-n+\nu}{2}, \operatorname{sgn} t\right),$$

$$(118)\quad H(y; S, u) = G(y; S, u) + \frac{m-2n}{4\pi} F(y; S, u),$$

$$F^*(y; S, u) = \alpha_0(S, u)\, W_0\left(y, \frac{m}{2}+\nu\right) + \beta_0(S, u) + F(y; S, u),$$

$$H^*(y; S, u) = G(y; S, u) + \frac{m-2n}{4\pi} F^*(y; S, u).$$

Die Beziehungen, die sich aus (113) ergeben, wenn man $x = 0$ setzt und die Ableitungen nach x an der Stelle $x = 0$ bildet, lassen sich in die Gestalt

$$(119)\quad \begin{aligned} F^*\left(\frac{1}{y}; S, u\right) &= \|S\|^{-\frac{1}{2}} y^{\frac{m}{2}+\nu} F^*(y; S^{-1}, \tilde{u}), \\ H^*\left(\frac{1}{y}; S, u\right) &= -\|S\|^{-\frac{1}{2}} y^{\frac{m}{2}+\nu} H^*(y; S^{-1}, \tilde{u}) \end{aligned}$$

bringen. Die ungestirnten Funktionen genügen daher den Transformationsformeln

$$\begin{aligned} F\left(\frac{1}{y}; S, u\right) &= \|S\|^{-\frac{1}{2}} y^{\frac{m}{2}+\nu} F(y; S^{-1}, \tilde{u}) + \\ &+ \|S\|^{-\frac{1}{2}} y^{\frac{m}{2}+\nu}\left\{\alpha_0(S^{-1}, \tilde{u})\, W_0\left(y, \frac{m}{2}+\nu\right) + \beta_0(S^{-1}, \tilde{u})\right\} - \\ &- \left\{\alpha_0(S, u)\, W_0\left(\frac{1}{y}, \frac{m}{2}+\nu\right) + \beta_0(S, u)\right\} \end{aligned}$$

und

$$\begin{aligned} H\left(\frac{1}{y}; S, u\right) &= -\|S\|^{-\frac{1}{2}} y^{\frac{m}{2}+\nu} H(y; S^{-1}, \tilde{u}) + \\ &+ \frac{2n-m}{4\pi} \|S\|^{-\frac{1}{2}} y^{\frac{m}{2}+\nu}\left\{\alpha_0(S^{-1}, \tilde{u})\, W_0\left(y, \frac{m}{2}+\nu\right) + \beta_0(S^{-1}, \tilde{u})\right\} + \\ &+ \frac{2n-m}{4\pi}\left\{\alpha_0(S, u)\, W_0\left(\frac{1}{y}, \frac{m}{2}+\nu\right) + \beta_0(S, u)\right\}. \end{aligned}$$

Für die Funktionen

$$(120)\quad \xi(s; S, u) = \int_0^\infty F(y; S, u)\, y^{s-1}\, dy, \quad \eta(s; S, u) = \int_0^\infty H(y; S, u)\, y^{s-1}\, dy$$

ergeben sich damit nach elementaren Integrationen in bekannter Weise die Darstellungen

$$\begin{aligned} \xi(s;\,S,u) &= \int_1^\infty \left\{ y^s F(y;\,S,u) + \|S\|^{-\frac{1}{2}}\, y^{\frac{m}{2}+\nu-s} F(y;\,S^{-1},\bar{u}) \right\} \frac{dy}{y} + \\ &+ \frac{\alpha_0(S,u)}{s\left(s+1-\frac{m}{2}-\nu\right)} + \|S\|^{-\frac{1}{2}} \frac{\alpha_0(S^{-1},\bar{u})}{(s-1)\left(s-\frac{m}{2}-\nu\right)} - \\ &- \frac{\beta_0(S,u)}{s} + \|S\|^{-\frac{1}{2}} \frac{\beta_0(S^{-1},\bar{u})}{s-\frac{m}{2}-\nu} \end{aligned} \tag{121}$$

und

$$\begin{aligned} \eta(s;\,S,u) &= \int_1^\infty \left\{ y^s H(y;\,S,u) - \|S\|^{-\frac{1}{2}}\, y^{\frac{m}{2}+\nu-s} H(y;\,S^{-1},\bar{u}) \right\} \frac{dy}{y} + \\ &+ \frac{2n-m}{4\pi} \left\{ \|S\|^{-\frac{1}{2}} \frac{\alpha_0(S^{-1},\bar{u})}{(s-1)\left(s-\frac{m}{2}-\nu\right)} - \frac{\alpha_0(S,u)}{s\left(s+1-\frac{m}{2}-\nu\right)} + \right. \\ &\left. + \|S\|^{-\frac{1}{2}} \frac{\beta_0(S^{-1},\bar{u})}{s-\frac{m}{2}-\nu} + \frac{\beta_0(S,u)}{s} \right\}. \end{aligned} \tag{122}$$

Andererseits erhält man durch gliedweise Integration unter Verwendung der in § 4 diskutierten Gammafaktoren, wenn man noch

$$W(y;\,\alpha,\beta,-1) = W(y;\,\beta,\alpha,1)$$

berücksichtigt,

$$\begin{aligned} \xi(s;\,S,u) &= \pi^{-s}\Gamma\left(s;\,\frac{n+\nu}{2},\,\frac{m-n+\nu}{2}\right)\varphi\left(s-\frac{\nu}{2};\,S,u\right) + \\ &+ \pi^{-s}\Gamma\left(s;\,\frac{m-n+\nu}{2},\,\frac{n+\nu}{2}\right)\psi\left(s-\frac{\nu}{2};\,S,u\right) \end{aligned} \tag{123}$$

und

$$\begin{aligned} \eta(s;\,S,u) &= \frac{m-2n}{4\pi}\,\xi(s;\,S,u) + \\ &+ \pi^{-s-1}\Gamma\left(s+1;\,\frac{n+\nu}{2},\,\frac{m-n+\nu}{2}\right)\varphi\left(s-\frac{\nu}{2};\,S,u\right) - \\ &- \pi^{-s-1}\Gamma\left(s+1;\,\frac{m-n+\nu}{2},\,\frac{n+\nu}{2}\right)\psi\left(s-\frac{\nu}{2};\,S,u\right) \end{aligned} \tag{124}$$

mit

$$\varphi(s;\,S,u) = \sum_{t>0} \alpha_t(S,u)\, t^{\frac{m}{2}-1-s}, \quad \psi(s;\,S,u) = \sum_{t<0} \alpha_t(S,u)\, |t|^{\frac{m}{2}-1-s}. \tag{125}$$

Der meromorphe Charakter der Funktionen $\xi(s;\,S,u)$ und $\eta(s;\,S,u)$ ist evident auf Grund der Darstellungen (121) und (122). Die Relationen (123) und (124) lassen sich nach φ und ψ auflösen, da die hier auftretende Nenner-

determinante $D\left(s; \frac{n+\nu}{2}, \frac{m-n+\nu}{2}\right)$ nicht verschwindet. Mit (75) ergibt sich

$$\varphi\left(s-\frac{\nu}{2}; S, u\right) = \frac{\pi^s}{2} \frac{\Gamma\left(s+1; \frac{m-n+\nu}{2}, \frac{n+\nu}{2}\right)}{\Gamma(s)\,\Gamma\left(s+1-\frac{m}{2}-\nu\right)} \xi(s; S, u) +$$

(126)

$$+\frac{\pi^{s+1}}{2} \frac{\Gamma\left(s; \frac{m-n+\nu}{2}, \frac{n+\nu}{2}\right)}{\Gamma(s)\,\Gamma\left(s+1-\frac{m}{2}-\nu\right)} \left\{\eta(s; S, u) + \frac{2n-m}{4\pi}\,\xi(s; S, u)\right\}$$

und

$$\psi\left(s-\frac{\nu}{2}; S, u\right) = \frac{\pi^s}{2} \frac{\Gamma\left(s+1; \frac{n+\nu}{2}, \frac{m-n+\nu}{2}\right)}{\Gamma(s)\,\Gamma\left(s+1-\frac{m}{2}-\nu\right)} \xi(s; S, u) -$$

(127)

$$-\frac{\pi^{s+1}}{2} \frac{\Gamma\left(s; \frac{n+\nu}{2}, \frac{m-n+\nu}{2}\right)}{\Gamma(s)\,\Gamma\left(s+1-\frac{m}{2}-\nu\right)} \left\{\eta(s; S, u) + \frac{2n-m}{4\pi}\,\xi(s; S, u)\right\}.$$

Man erkennt nun, daß auch die Funktionen $\varphi(s; S, u)$ und $\psi(s; S, u)$ meromorph sind.

Die Formel (71) zeigt unmittelbar, daß $\Gamma(s; \alpha, \beta)$ in der Halbebene

$$\sigma = \Re s > \mathrm{Max}(0, \Re(\alpha+\beta)-1)$$

regulär ist. Aus (89) entnehmen wir noch

$$\nu = \frac{1}{2}\left(2-m+\sqrt{(m-2)^2-4\lambda}\right), \tag{128}$$

wobei die Wurzel eine nicht-negative reelle Zahl ist oder positiven Imaginärteil hat, je nachdem $0 \leq \lambda \leq \frac{m-2}{2}$ oder $\lambda > \frac{m-2}{2}$ gilt. Mithin sind $\varphi\left(s-\frac{\nu}{2}; S, u\right)$ und $\psi\left(s-\frac{\nu}{2}; S, u\right)$ in der Halbebene

$$\sigma > \mathrm{Max}\left(0, \frac{m}{2}+\Re\nu-1\right) = \begin{cases} \frac{1}{2}\sqrt{(m-2)^2-4\lambda} & \text{für } 0 \leq \lambda \leq \frac{m-2}{2}, \\ 0 & \text{für } \lambda \geq \frac{m-2}{2} \end{cases} \tag{129}$$

höchstens dort singulär, wo $\xi(s; S, u)$ und $\eta(s; S, u)$ singulär sind. Nach (121) und (122) kann es sich höchstens um die Stellen 1 und $\frac{m}{2}+\nu$ in der s-Ebene handeln. Hieraus schließt man, daß $\varphi(s; S, u)$ und $\psi(s; S, u)$ in der Halbebene

$$\sigma > \Re\frac{m+\nu}{2} = \begin{cases} \frac{1}{4}\left(m+2+\sqrt{(m-2)^2-4\lambda}\right) & \text{für } 0 \leq \lambda \leq \frac{m-2}{2}, \\ \frac{1}{4}(m+2) & \text{für } \lambda \geq \frac{m-2}{2} \end{cases} \tag{130}$$

regulär sind. Es kann also zu jedem Eigenwert $\lambda \geq 0$ eine positive Zahl $\delta = \delta(\lambda)$ so bestimmt werden, daß für $m \geq 3$ die Funktionen $\varphi(s; S, u)$ und $\psi(s; S, u)$ in der Halbebene $\sigma > \frac{m}{2}-\delta$ regulär sind mit Ausnahme des Punktes $s = \frac{m}{2}$

falls $\lambda = 0$ ist. In diesem Fall sind noch die Residuen von Interesse. Sie lauten

$$\operatorname*{Res}_{s=\frac{m}{2}} \varphi(s; S, u) = \frac{\pi^{\frac{m}{2}}}{\Gamma\left(\frac{m}{2}\right)} \|S\|^{-\frac{1}{2}} \left(\frac{\alpha_0(S^{-1}, \bar{u})}{m-2} + \frac{\beta_0(S^{-1}, \bar{u})}{2}\right) \gamma_1 ,$$

$$\operatorname*{Res}_{s=\frac{m}{2}} \psi(s; S, u) = \frac{\pi^{\frac{m}{2}}}{\Gamma\left(\frac{m}{2}\right)} \|S\|^{-\frac{1}{2}} \left(\frac{\alpha_0(S^{-1}, \bar{u})}{m-2} + \frac{\beta_0(S^{-1}, \bar{u})}{2}\right) \gamma_2$$

mit

$$\gamma_1 = \Gamma\left(\frac{m}{2} + 1; \frac{m-n}{2}, \frac{n}{2}\right) + \frac{2n-m}{2} \Gamma\left(\frac{m}{2}; \frac{m-n}{2}, \frac{n}{2}\right),$$
$$\gamma_2 = \Gamma\left(\frac{m}{2} + 1; \frac{n}{2}, \frac{m-n}{2}\right) + \frac{m-2n}{2} \Gamma\left(\frac{m}{2}; \frac{n}{2}, \frac{m-n}{2}\right).$$

Allgemein handelt es sich bei diesen Zahlwerten um den Wert der Funktion

$$\Gamma(s + 2; \alpha, \beta) + (\beta - \alpha)\, \Gamma(s + 1; \alpha, \beta),$$

welche mit Hilfe von (71) und (72) in

$$s(s + 1 - \alpha - \beta)\, \Gamma(s; \alpha, \beta)$$
$$= 2^{\frac{\alpha+\beta}{2} - s} \frac{\Gamma(s+1)\, \Gamma(s+2-\alpha-\beta)}{\Gamma(s+1-\alpha)} F\left(s, s + 1 - \alpha - \beta; s + 1 - \alpha; \frac{1}{2}\right)$$

übergeführt wird, an der Stelle $s = \alpha + \beta - 1$. Man erhält

$$\Gamma(\alpha + \beta + 1; \alpha, \beta) + (\beta - \alpha)\, \Gamma(\alpha + \beta; \alpha, \beta) = 2^{1 - \frac{\alpha+\beta}{2}} \frac{\Gamma(\alpha+\beta)}{\Gamma(\beta)},$$

damit auch

$$\operatorname*{Res}_{s=\frac{m}{2}} \varphi(s; S, u) = 2^{1-\frac{m}{4}} \frac{\pi^{\frac{m}{2}}}{\Gamma\left(\frac{n}{2}\right)} \|S\|^{-\frac{1}{2}} \left(\frac{\alpha_0(S^{-1}, \bar{u})}{m-2} + \frac{\beta_0(S^{-1}, \bar{u})}{2}\right), \tag{131}$$

$$\operatorname*{Res}_{s=\frac{m}{2}} \psi(s; S, u) = 2^{1-\frac{m}{4}} \frac{\pi^{\frac{m}{2}}}{\Gamma\left(\frac{m-n}{2}\right)} \|S\|^{-\frac{1}{2}} \left(\frac{\alpha_0(S^{-1}, \bar{u})}{m-2} + \frac{\beta_0(S^{-1}, \bar{u})}{2}\right), \tag{132}$$

wobei $\lambda = 0$ vorausgesetzt wurde.

Bezüglich des Verhaltens in einer Halbebene $\sigma \geqq \sigma_0$ ist mit den üblichen Schlüssen zu zeigen, daß $\varphi(s; S, u)$ und $\psi(s; S, u)$ für $|\mathfrak{Im}\, s| \to \infty$ gleichmäßig in σ höchstens wie $e^{C|\mathfrak{Im}\, s|}$ wachsen, wobei C eine positive Konstante ist.

§ 8. Verteilung von Gitterpunkten im ternären Fall

Die Voraussetzungen von § 6 bezüglich S werden wieder eingeführt und bis zum Schluß beibehalten. Nun ist, wie in § 5 ausgeführt wurde,

$$\alpha_0(S, u) = \alpha_0(S^{-1}, \bar{u}) = 0, \quad \beta_0(S, u) = \beta_0(S^{-1}, \bar{u}) = I\{u\},$$

also

$$\begin{aligned} &\operatorname*{Res}_{s=\frac{3}{2}} \varphi(s; S, u) = 2^{-\frac{1}{4}} \pi \|S\|^{-\frac{1}{2}} I\{u\}, \\ &\operatorname*{Res}_{s=\frac{3}{2}} \psi(s; S, u) = 2^{-\frac{1}{4}} \pi^{\frac{3}{2}} \|S\|^{-\frac{1}{2}} I\{u\}, \end{aligned} \tag{133}$$

außerdem den Formeln (105), (107), (108), (116), (117), (125), (131), (132) zufolge

$$\varphi(s; S, u) = \frac{2^{-\frac{3}{2}}\pi}{V(S)} \sum_{\substack{\mathfrak{g}/\Gamma(S)\\ S[\mathfrak{g}]>0}} \frac{u(\mathfrak{g})}{E(S,\mathfrak{g})} (S[\mathfrak{g}])^{-s},$$

(134)

$$\psi(s; S, u) = \frac{2^{-\frac{3}{2}}\sqrt{\pi}}{V(S)} \sum_{\substack{\mathfrak{g}/\Gamma(S)\\ S[\mathfrak{g}]<0}} l(S,\mathfrak{g})\, u(\mathfrak{g})\, |S[\mathfrak{g}]|^{-s}.$$

Die Gleichungen (133) sind allgemein gültig, da im Falle $\lambda > 0$ der Mittelwert $I\{u\}$ verschwindet.

Da $\Gamma^*(S)$ Grenzkreisgruppe erster Art ohne parabolische Substitutionen ist, so hat Δ_S, wie W. Roelcke [7] gezeigt hat, ein diskretes Spektrum; d. h. die Eigenwerte $\lambda(\geqq 0)$ haben endliche Vielfachheit und häufen sich im Endlichen nicht. Insbesondere hat $\lambda = 0$ die Vielfachheit 1; die zugehörige Eigenfunktion ist konstant. Ferner läßt sich jede zweimal stetig differenzierbare Funktion $f(\tau)$, die gegenüber $\Gamma^*(S)$ invariant ist, in eine gleichmäßig konvergente Reihe nach den Eigenfunktionen $u(\tau)$ entwickeln und jede stetige, bezüglich $\Gamma^*(S)$ invariante Funktion durch endliche Linearkombinationen der $u(\tau)$ gleichmäßig approximieren.

Es sei $\mathfrak{B}^*$ ein beliebiger, im Riemannschen Sinne meßbarer Teilbereich des kompakten und im gleichen Sinne meßbaren Fundamentalbereichs $\mathfrak{F}^*(S)$, den wir uns als Normalpolygon gewählt denken können. $\mathfrak{B}_0$ und $\mathfrak{F}_0(S)$ seien die entsprechenden Bereiche in $S[\mathfrak{x}] > 0$, die mit $\mathfrak{x}$ auch $r\mathfrak{x}$ $(r > 0)$ enthalten mögen. Das Ziel der folgenden Überlegungen ist der Nachweis der asymptotischen Relation

$$\sum_{\substack{\mathfrak{g}\in\mathfrak{B}_0\\ 0<S[\mathfrak{g}]\leq q}} \frac{1}{E(S,\mathfrak{g})} \sim \frac{2\pi}{3} \|S\|^{-\frac{1}{2}} V(\mathfrak{B}_0)\, q^{\frac{3}{2}} \quad \text{für } q\to\infty. \tag{135}$$

Dabei ist $V(\mathfrak{B}_0) = V(\mathfrak{B}^*)$ der hyperbolische Inhalt von $\mathfrak{B}^*$. Ist $f(\mathfrak{x}) = f(\tau)$ die charakteristische Funktion von $\mathfrak{B}^*$:

$$f(\tau) = \begin{cases} 1 & \text{für } \tau\in\mathfrak{B}^*,\\ 0 & \text{für } \tau\in\mathfrak{F}^*(S)-\mathfrak{B}^* \end{cases}$$

und ist $f(\tau)$ überdies gegenüber $\Gamma^*(S)$ invariant, so kann (135) in der Form

$$\sum_{\substack{\mathfrak{g}\in\mathfrak{F}_0(S)\\ 0<S[\mathfrak{g}]\leq q}} \frac{f(\mathfrak{g})}{E(S,\mathfrak{g})} \sim \frac{2\pi}{3} \|S\|^{-\frac{1}{2}} (f,1)\, q^{\frac{3}{2}} \quad \text{für } q\to\infty \tag{136}$$

geschrieben werden, wenn allgemein unter (f, g) das Skalarprodukt

$$(f,g) = \int_{\mathfrak{F}^*(S)} f(\tau)\, g(\tau)\, dv$$

zweier reeller, bezüglich $\Gamma^*(S)$ invarianter Funktionen $f(\tau), g(\tau)$ verstanden wird. Der Beweis von (136) kann wegen der Linearität dieser Beziehung in f mit den in [4] angegebenen Schlüssen erbracht werden, indem man die charakteristische Funktion $f(\mathfrak{x})$ zunächst in gewisser Weise durch eine stetige

Funktion approximiert, diese wiederum durch endliche Linearkombinationen der Eigenfunktionen $u(\mathfrak{x})$, so daß (136) nur noch für eine Eigenfunktion $u(\mathfrak{x})$ an Stelle von $f(\mathfrak{x})$ geprüft zu werden braucht. Diesen Spezialfall beweist man mit Hilfe eines allgemeinen Tauberschen Satzes (Hardy-Littlewood), der auf $\varphi(s; S, u)$ angewendet wird. Daß $\varphi(s; S, u)$ die Voraussetzungen dieses Satzes erfüllt, wurde in § 7 festgestellt. Schließlich ist noch zu bemerken, daß es in $\mathfrak{F}^*(S)$ nur endlich viele elliptische Fixpunkte gibt, so daß $E(S, \mathfrak{g})$ meist 1 ist und (135) auch gültig bleibt, wenn man $E(S, \mathfrak{g})$ durch 1 ersetzt. So wird man auf die eingangs genannte asymptotische Gitterpunktsformel (27) geführt.

Literatur

[1] *Bateman Manuscript Project:* Tables of Integral Transforms. Vol. I. New York 1954. — [2] Fricke, R., u. F. Klein: Vorlesungen über die Theorie der automorphen Funktionen. I. Leipzig 1897. — [3] Huber, H.: Über eine neue Klasse automorpher Funktionen und ein Gitterpunktsproblem in der hyperbolischen Ebene. Comm. Math. Helvetici **30**, 20—62 (1956). — [4] Maass, H.: Über die Verteilung der zweidimensionalen Untergitter in einem euklidischen Gitter. Math. Ann. **137**, 319—327 (1959). — [5] Maass, H.: Zur Theorie der Kugelfunktionen einer Matrixvariablen. Math. Ann. **135**, 319—416 (1958). — [6] Maass, H.: Die Differentialgleichungen in der Theorie der elliptischen Modulfunktionen. Math. Ann. **125**, 235—263 (1953). — [7] Roelcke, W.: Über die Wellengleichung bei Grenzkreisgruppen erster Art. Sitzungsber. Heidelberger Akad. Wiss., Math.-naturwiss. Kl. **1953/55**, 4. Abh. 161—267. — [8] Siegel, C. L.: Indefinite quadratische Formen und Funktionentheorie. I. Math. Ann. **124**, 17—54 (1951).

(Eingegangen am 7. Juni 1959)

Kerner, H.
Math. Annalen 138, 316—328 (1959)

Holomorphiehüllen zu K-vollständigen komplexen Räumen

Von
Hans Kerner in München

Einleitung

Im Jahre 1932 hat P. Thullen [19] gezeigt, daß man jedem unverzweigten Riemannschen Gebiet eine unverzweigte Holomorphiehülle zuordnen kann. Der Begriff der Holomorphiehülle stellt für die Behandlung vieler Probleme ein wichtiges Hilfsmittel dar. Die Entwicklung der Funktionentheorie wie auch der Topologie hat in jüngster Zeit dazu geführt, auch verzweigte Riemannsche Gebiete und, noch allgemeiner, komplexe Räume zu betrachten. Es ist daher erwünscht, für allgemeinere Gebilde, als es die unverzweigten Riemannschen Gebiete sind, Holomorphiehüllen zu definieren und deren Existenz nachzuweisen.

Wir zeigen in § 1, daß man jedem verzweigten Riemannschen Gebiet (X, Φ, C^n) eine Holomorphiehülle zuordnen kann (Satz 2)[1]. Man ersieht jedoch an Beispielen, die bereits 1957 von H. Grauert und R. Remmert ([9], [10]) angegeben wurden, daß diese Holomorphiehülle im allgemeinen nicht invariant gegenüber biholomorphen Abbildungen ist. Sie hängt vielmehr von der Projektionsabbildung Φ ab, durch die der komplexe Raum X über dem C^n konkretisiert wird. Wir nennen daher diese Holomorphiehülle die Φ-Hülle. Nun ist bekannt, daß die von P. Thullen definierte unverzweigte Hülle eines unverzweigten Gebiets gegenüber biholomorphen Abbildungen invariant ist (H. Cartan [4], VII, théorème 2). Von diesem Standpunkt aus betrachtet, stellen also die Φ-Hüllen noch nicht das genaue Analogon zu den unverzweigten Hüllen dar. Wir müssen uns daher die Frage stellen, ob man jedem verzweigten Riemannschen Gebiet eine von der speziellen Konkretisierung unabhängige Hülle zuordnen kann. Dieses Problem wird in § 2 behandelt und die Frage im positiven Sinne beantwortet.

Wir zeigen in § 2, daß es möglich ist, jedem K-vollständigen komplexen Raum eine Holomorphiehülle zuzuordnen (Satz 3). Diese Hülle nennen wir die K-Hülle. Sie ist invariant gegenüber biholomorphen Abbildungen.

In § 3 untersuchen wir Beziehungen zwischen Φ-Hülle und K-Hülle. Ist (X, Φ, C^n) ein Riemannsches Gebiet, so ist X ein K-vollständiger komplexer Raum. Man kann daher die Φ-Hülle und die K-Hülle bilden. Die Φ-Hülle ist als offenes dichtes Teilgebiet in der K-Hülle enthalten. Die beiden Hüllen unterscheiden sich genau um die Entartungsmenge der in die K-Hülle fortgesetzten

[1]) Diese Aussage wurde auch von R. Iwahashi [12] und G. Scheja [17] bewiesen.

Abbildung Φ (Satz 4). Wenn die Φ-Hülle von der K-Hülle verschieden ist, so ist die Φ-Hülle sicher nicht holomorph-konvex (Satz 6). (Vgl. dazu die von H. Grauert und R. Remmert in [9] angegebenen Beispiele.) Für unverzweigte Gebiete stimmen Φ-Hülle und K-Hülle mit der von P. Thullen definierten unverzweigten Hülle überein (Satz 8).

Es sei mir gestattet, an dieser Stelle Herrn Professor Dr. K. Stein für wertvolle Anregungen und Ratschläge zu danken.

§ 0. Vorbereitende Bemerkungen

Die im folgenden verwendeten topologischen Begriffe wie „eigentliche Abbildung", „nirgends zerlegende Menge" usw. seien wie in [3] und [18] erklärt. Unter einem Filter $\mathfrak{F}$ auf einem Raum R wollen wir in dieser Arbeit ein nichtleeres System von nichtleeren Teilmengen von R verstehen, das die Eigenschaft besitzt: Zu $U, V \in \mathfrak{F}$ existiert immer ein $W \in \mathfrak{F}$ mit $W \subset U \cap V$.

Eine Abbildung $\varphi: R_1 \to R_2$ eines zusammenhängenden topologischen Raumes R_1 in einen topologischen Raum R_2 heiße *nirgends entartet*, wenn für jeden Punkt $p_2 \in R_2$ die Faser $\overset{-1}{\varphi}(p_2) \subset R_1$ höchstens aus isolierten Punkten besteht und $\varphi(R_1)$ eine offene Menge von R_2 enthält.

Wir beweisen einen Hilfssatz, den wir später mehrmals anwenden werden:

Hilfssatz 1: *Es seien X und Y lokal-kompakte lokal-zusammenhängende Hausdorffsche Räume und $\varphi: X \to Y$ eine stetige nirgends entartete Abbildung von X in Y. Ist dann $\mathfrak{F}$ ein Filter von zusammenhängenden Teilmengen von X und konvergiert der Filter $\varphi(\mathfrak{F})$, so gilt: Entweder besitzt $\mathfrak{F}$ keinen Berührpunkt in X oder $\mathfrak{F}$ konvergiert gegen einen Punkt von X.*

Beweis: Es sei $x \in X$ ein Berührpunkt von $\mathfrak{F}$. Wir zeigen, daß dann $\mathfrak{F}$ gegen x konvergiert. Zu jeder beliebigen Umgebung A von x gibt es eine Umgebung U von x, $U \subset A$, mit folgender Eigenschaft: $\varphi \mid U$[2]) ist eine eigentliche Abbildung von U in eine Umgebung W von $\varphi(x)$ (Vgl. K. Stein [18], Hilfssatz 3). Es sei W_1 eine Umgebung von $\varphi(x)$, die relativ-kompakt in W liegt. Dann ist $U_1 := \overset{-1}{\varphi}(W_1) \cap U$ eine Umgebung von x mit $U_1 \Subset U$. $\varphi(\mathfrak{F})$ konvergiert gegen $\varphi(x)$. Daher gibt es ein $F \in \mathfrak{F}$ mit $\varphi(F) \subset W_1$; also ist $F \subset \overset{-1}{\varphi}(\varphi(F)) \subset \overset{-1}{\varphi}(W_1)$. Die Menge $F \cap U =: F_1$ ist nicht leer, weil x ein Berührpunkt von $\mathfrak{F}$ ist. Wir betrachten nun F als Teilraum von X, versehen mit der induzierten Topologie. Bezüglich dieser Topologie ist F_1 eine offene Teilmenge von F. F_1 ist auch eine abgeschlossene Teilmenge von F, denn jeder Häufungspunkt p von F_1, der in F liegt, gehört wegen $F_1 \subset U_1 \Subset U$ zu U; daher gehört p zu $F \cap U = F_1$. Also ist F_1 eine nichtleere offene und abgeschlossene Teilmenge von F. Nach Voraussetzung ist F zusammenhängend; somit ist $F = F_1$, also $F \subset U \subset A$. Das bedeutet, daß es zu jeder Umgebung A von x ein $F \in \mathfrak{F}$ gibt mit $F \subset A$, d. h. $\mathfrak{F}$ konvergiert gegen x.

Unter einem komplexen Raum wollen wir in dieser Arbeit immer einen zusammenhängenden komplexen Raum im Sinne der Definition von H. Behnke und K. Stein verstehen, der nach einem Satz von H. Grauert und

[2]) $\varphi|U$ bedeutet die Beschränkung der Abbildung φ auf U.

R. REMMERT ([11], Satz 33) auch ein (normaler) komplexer Raum nach der Definition von H. CARTAN ist. Zur Definition der damit in Verbindung stehenden Begriffe, insbesondere der analytisch-verzweigten Überlagerung eines Polyzylinders, verweisen wir auf [18] und [8].

Die Menge aller in einem komplexen Raum X holomorphen Funktionen bildet einen (Integritäts-) Ring, den wir mit $\mathfrak{R}(X)$ bezeichnen. Es gilt:

Hilfssatz 2: *Es sei $\varphi: X \to Y$ eine eigentliche nirgends entartete holomorphe Abbildung eines komplexen Raumes X auf einen komplexen Raum Y. Dann ist $\mathfrak{R}(X)$ algebraisch über $\mathfrak{R}(Y)$ vermöge φ, d. h. zu jeder Funktion $f \in \mathfrak{R}(X)$ gibt es Funktionen $c_\sigma \in \mathfrak{R}(Y)$, $\sigma = 1, \ldots, s$ (s nat. Z.), so daß gilt:*

$$f^s(x) + c_1(\varphi(x)) \cdot f^{s-1}(x) + \cdots + c_s(\varphi(x)) = 0\,, \quad x \in X\,.$$

Zum Beweis vgl. etwa [11], Satz 12.

Wir benötigen ferner:

Hilfssatz 3: *Es seien X, Y, Z komplexe Räume, $f: X \to Y$ eine holomorphe Abbildung von X in Y und $\varphi: X \to Z$ eine eigentliche lokal-topologische holomorphe Abbildung von X auf Z. $S(f)$ sei die Menge aller Punkte $z \in Z$ mit folgender Eigenschaft: Zu z gibt es Punkte $x_1, x_2 \in X$, $x_1 \neq x_2$, mit $f(x_1) = f(x_2)$ und $\varphi(x_1) = \varphi(x_2) = z$. Dann ist $S(f)$ eine analytische Menge in Z.*

Beweis: 1) $S(f)$ ist abgeschlossen in Z. — Ist $z \in Z - S(f)$ und sind $x_1, \ldots, x_s$ die Punkte von $\overset{-1}{\varphi}(z)$, dann sind die Punkte $f(x_1), \ldots, f(x_s)$ paarweise verschieden. Es gibt paarweise disjunkte Umgebungen $V(f(x_1)), \ldots, V(f(x_s))$. Dann ist $\bigcup_{\sigma=1}^{s} \overset{-1}{f}\big(V(f(x_\sigma))\big) =: U$ eine Umgebung von $\overset{-1}{\varphi}(z)$. Es gibt dazu eine Umgebung W von z mit $\overset{-1}{\varphi}(W) \subset U$ (vgl. K. STEIN [18], Hilfssatz 1); offenbar ist $W \cap S(f) = \varnothing$. 2) Ist $z_0 \in S(f)$, so sei jeweils U_σ eine schlichte Umgebung von $x_\sigma (\sigma = 1, \ldots, s)$; die Abbildungen $\varphi_\sigma := \varphi \mid U_\sigma$ sind dann biholomorph. In einer Umgebung W von z_0 wird $S(f)$ gegeben als Vereinigung des Nullstellenbildes endlich vieler Gleichungen $f \circ \varphi_\sigma^{-1} = f \circ \varphi_\tau^{-1}$ $(\sigma, \tau = 1, \ldots, s)$. Eine Umgebung von $f \circ \varphi_\sigma^{-1}(z_0)$ wird durch lokal holomorphe Funktionen $h_\varkappa (\varkappa = 1, \ldots, k)$ separiert. Die Gleichungen $f \circ \varphi_\sigma^{-1} = f \circ \varphi_\tau^{-1}$ sind dann in einer Umgebung von z_0 äquivalent mit den Gleichungen $h_\varkappa \circ f \circ \varphi_\sigma^{-1} - h_\varkappa \circ f \circ \varphi_\tau^{-1} = 0$.

Sind X und B komplexe Räume, so heißt ein Tripel (X, Φ, B) ein *Gebiet über B*, wenn $\Phi: X \to B$ eine nirgends entartete holomorphe Abbildung von X in B ist. Ist B der Raum C^n der n komplexen Veränderlichen $z_1, \ldots, z_n$, so heißt (X, Φ, C^n) ein *Riemannsches Gebiet*. Man definiert (vgl. [9]):

Definition 1: *Ein Randpunkt eines Gebietes (X, Φ, B) ist ein Filter r auf X mit folgenden Eigenschaften:*

a) *r ist ein Filter von offenen zusammenhängenden Teilmengen von X.*

b) *r besitzt in X keinen Berührpunkt.*

c) *$\Phi(r)$ konvergiert gegen einen Punkt $b_0 \in B$.*

d) *Ist $A \subset B$ eine offene Umgebung von b_0, so gehört genau eine zusammenhängende Komponente von $\overset{-1}{\Phi}(A)$ zu r. r wird durch diese Mengen erzeugt.*

Die Menge aller Randpunkte von (X, Φ, B) bezeichnen wir kurz mit $Rd\,X$. In $\overline{X} := X \cup Rd\,X$ führt man eine Topologie ein: Eine Umgebung von $r_0 \in Rd\,X$ ist eine Menge $U \in r_0$ vereinigt mit allen $r \in Rd\,X$, zu denen ein $V \in r$ mit $V \subset U$ existiert. $\overline{X}$ ist ein Hausdorffscher Raum. Durch die Festsetzung $\overline{\Phi}(x) := \Phi(x)$ für $x \in X$ und $\overline{\Phi}(r) := b_0$ für $r \in Rd\,X$ wird Φ zu einer stetigen Abbildung $\overline{\Phi} : \overline{X} \to B$ fortgesetzt.

Aus Forderung a) in Definition 1 ergibt sich unmittelbar, daß $Rd\,X$ nirgends zerlegend in $\overline{X}$ ist.

§ 1. Holomorphiehüllen zu verzweigten Riemannschen Gebieten

In diesem Paragraphen wollen wir zu Riemannschen Gebieten, die Verzweigungspunkte als innere Punkte enthalten dürfen, Holomorphiehüllen konstruieren.

Sind X und Y komplexe Räume und $\mathfrak{R}(X)$ bzw. $\mathfrak{R}(Y)$ die Ringe der in X bzw. Y holomorphen Funktionen, so induziert jede holomorphe Abbildung $\varphi : X \to Y$ einen Homomorphismus $^*\varphi : \mathfrak{R}(Y) \to \mathfrak{R}(X)$, wenn man definiert: $^*\varphi(f) := f \circ \varphi$, $f \in \mathfrak{R}(Y)$.

Nun erklären wir:

Definition 2: *Ein Riemannsches Gebiet $(\hat{X}, \hat{\Phi}, C^n)$ heißt die Φ-Hülle des Riemannschen Gebiets (X, Φ, C^n), wenn gilt:*

a) *Es gibt eine nirgends entartete holomorphe Abbildung $\varphi : X \to \hat{X}$ mit $\Phi = \hat{\Phi} \circ \varphi$, so daß der durch φ induzierte Homomorphismus $^*\varphi : \mathfrak{R}(\hat{X}) \to \mathfrak{R}(X)$ ein Isomorphismus von $\mathfrak{R}(\hat{X})$ auf $\mathfrak{R}(X)$ ist.*

b) *Zu jedem Riemannschen Gebiet (Y, Ψ, C^n), zu dem es eine nirgends entartete holomorphe Abbildung $\psi : X \to Y$ mit $\Phi = \Psi \circ \psi$ gibt, so daß $^*\psi : \mathfrak{R}(Y) \to \mathfrak{R}(X)$ ein Isomorphismus von $\mathfrak{R}(Y)$ auf $\mathfrak{R}(X)$ ist, existiert eine nirgends entartete holomorphe Abbildung $\tau : Y \to \hat{X}$, so daß gilt: $\varphi = \tau \circ \psi$.*

Für die Φ-Hülle schreiben wir kurz $\hat{H}(X, \Phi, C^n)$.

Grundlegend für die Bildung unverzweigter Hüllen ist ein Satz von H. Cartan ([4], VII, théorème 1; vgl. auch [2], [5] und [19]), der sich auf unverzweigte Gebiete beschränkt. Wir beweisen die entsprechende Aussage für beliebige (verzweigte) Gebiete.

Satz 1[3]**:** *Es sei (X, Φ, B) ein Gebiet über einem komplexen Raum B und F eine Menge holomorpher Abbildungen $f : X \to Y_f$ von X in komplexe Räume Y_f. Dann gibt es ein Gebiet $(\hat{X}, \hat{\Phi}, B)$ mit folgenden Eigenschaften:*

a) *Es gibt eine nirgends entartete holomorphe Abbildung $\varphi : X \to \hat{X}$ mit $\Phi = \hat{\Phi} \circ \varphi$, so daß jede Abbildung $f \in F$ nach $\hat{X}$ fortsetzbar ist.*

b) *Ist (Y, Ψ, B) ein Gebiet und gibt es eine nirgends entartete holomorphe Abbildung $\psi : X \to Y$ mit $\Phi = \Psi \circ \psi$, so daß jede Abbildung $f \in F$ nach Y fortsetzbar ist, so existiert eine holomorphe Abbildung $\tau : Y \to \hat{X}$ mit $\varphi = \tau \circ \psi$.*

Für das Gebiet $(\hat{X}, \hat{\Phi}, B)$ schreiben wir $\hat{H}(X, \Phi, B; F)$.

Beweis: Wir führen den Beweis in zwei Schritten. Zuerst wird der Raum $\hat{X}$ konstruiert und mit einer komplexen Struktur versehen. Dann zeigen wir, daß $\hat{X}$ die geforderten Eigenschaften besitzt.

[3]) Vgl. [1]).

1. Es sei X_0 das Teilgebiet von X, auf dem Φ lokal-topologisch abbildet. $(X_0, \Phi \mid X_0, B)$ ist ein unverzweigtes Gebiet; die Abbildungen $f \in F$ sind in X_0 holomorph. Nach dem Satz von H. Cartan ([4], VII, théorème 1) existiert dazu das maximale unverzweigte Gebiet $(\tilde{X}_0, \tilde{\Phi}_0, B)$, in das die Abbildungen aus F fortgesetzt werden können zu einer Menge von Abbildungen $\tilde{F}_0$. Nun sei $\tilde{X}_1$ die Menge aller Filter r auf $\tilde{X}_0$, die einen Randpunkt von $(\tilde{X}_0, \tilde{\Phi}_0, B)$ definieren und für die gilt: Für jede Abbildung $\tilde{f}_0 \in \tilde{F}_0$ konvergiert der Filter $\tilde{f}_0(r)$. Dann ist $\tilde{X}_2 := \tilde{X}_0 \cup \tilde{X}_1$ als Teilmenge von $\overline{\overline{X}}_0$ mit einer Topologie versehen. Die Abbildung $\tilde{\Phi}_0$ ist zu einer Abbildung $\tilde{\Phi} : \tilde{X}_2 \to B$ stetig fortsetzbar.

Es sei $\tilde{X}$ die Menge aller Punkte $\tilde{x} \in \tilde{X}_2$, die folgende Bedingung $(\tilde{U})$ erfüllen:

$(\tilde{U})$ Zu $\tilde{x}$ existiert eine Umgebung $\tilde{U}(\tilde{x}) \subset \tilde{X}_2$, so daß $\tilde{\Phi} \mid \tilde{U}(\tilde{x})$ eine nirgends entartete eigentliche Abbildung von $\tilde{U}(\tilde{x})$ auf eine offene Menge $Z \subset B$ ist.

$\tilde{X}$ ist, versehen mit der induzierten Topologie, ein lokalkompakter Hausdorffscher Raum. Jede Abbildung $\tilde{f}_0 \in \tilde{F}_0$ ist stetig fortsetzbar zu einer Abbildung $\tilde{f} : \tilde{X} \to Y_f$. Die Menge dieser Abbildungen sei mit $\tilde{F}$ bezeichnet.

Wir führen nun in $\tilde{X}$ eine komplexe Struktur ein.

Zu jedem Punkt $\tilde{x} \in \tilde{X}$ gibt es eine Umgebung $\tilde{U}$, die durch $\tilde{\Phi} \mid \tilde{U}$ eigentlich und nirgends entartet auf eine offene Menge $Z \subset B$ abgebildet wird. Wir dürfen annehmen, daß Z zu einer komplexen Karte $(Z, \mu, R, \varphi, E, M)$ von B gehört, d. h. (R, φ, E, M) ist eine analytisch-verzweigte Überlagerung eines Polyzylinders $E \subset C^n$ mit Verzweigungspunkten höchstens über der dünnen analytischen Menge $M \subset E$ und $\mu : Z \to R$ ist eine topologische Abbildung von Z auf R[4]). Außerdem können wir voraussetzen, daß $\overset{-1}{\tilde{\Phi}}(\tilde{\Phi}(\tilde{x})) \cap \tilde{U} = \{\tilde{x}\}$ ist. Dann ist $\lambda := \varphi \circ \mu \circ \tilde{\Phi}$ eine stetige eigentliche nirgends entartete Abbildung von $\tilde{U}$ auf E. Die Menge $\tilde{N} := \tilde{U} \cap (\tilde{X} - \tilde{X}_0)$ zerlegt nirgends in $\tilde{U}$. Dann zerlegt $M' := \lambda(\tilde{N})$ nirgends in E und das gleiche gilt für $M \cup M'$. Daher ist $\overset{-1}{\lambda}(M \cup M')$ nirgends zerlegend in $\tilde{U}$. Die Abbildung λ ist auf $\tilde{U} - \overset{-1}{\lambda}(M \cup M')$ lokaltopologisch. Wenn wir noch eine analytische Menge N in Z finden können mit $N \neq Z$ und $\tilde{\Phi}(\tilde{N}) \subset N$, dann gilt: λ ist eine lokaltopologische Abbildung von $\tilde{U} - \overset{-1}{\lambda}\big((\varphi \circ \mu(N)) \cup M\big)$ auf $E - (\varphi \circ \mu(N)) \cup M$. $\varphi \circ \mu$ ist eine eigentliche holomorphe Abbildung von Z auf E; nach einem Satz von R. Remmert ([16], Satz 23) ist $\varphi \circ \mu(N)$ eine analytische Menge in E. Dann ist auch $(\varphi \circ \mu(N)) \cup M$ eine analytische Menge in E und $(\tilde{U}, \lambda, E, (\varphi \circ \mu(N)) \cup M)$ ist eine analytisch-verzweigte Überlagerung von E mit Verzweigungspunkten höchstens über $(\varphi \circ \mu(N)) \cup M$.

Wir haben also nur noch eine analytische Menge N in Z mit $N \neq Z$ und $\tilde{\Phi}(\tilde{N}) \subset N$ zu konstruieren. Dazu benötigen wir eine Funktion g, die eine Faser $\overset{-1}{\tilde{\Phi}}(z_0) \cap \tilde{U}$, $z_0 \in Z - \tilde{\Phi}(\tilde{N})$, separiert.

Es sei $z \in Z - \tilde{\Phi}(\tilde{N})$ und $\{x_1(z), \ldots, x_s(z)\} = \overset{-1}{\tilde{\Phi}}(z) \cap \tilde{U}$ [5]) seien die über z liegenden Punkte von $\tilde{U}$. Nach Konstruktion von $\tilde{X}_0$ (s. H. Cartan [4]) gibt

[4]) Vgl. K. Stein [18] sowie H. Grauert und R. Remmert [8].

[5]) Die Abbildung $\tilde{\Phi} | \tilde{U} - \tilde{N}$ ist lokaltopologisch. Daher sind in einer Umgebung von z die $x_\sigma(z)$ als Umkehrabbildungen von $\tilde{\Phi}$ erklärt.

es zu $x_1(z)$, $x_2(z)$ eine Abbildung $\tilde{f}_{12} \in \tilde{F}$, die in diesen Punkten verschiedene Keime besitzt. Daher gibt es in beliebiger Nähe von z ein $z' \in Z - \tilde{\Phi}(\tilde{N})$, so daß $\tilde{f}_{12}(x_1(z')) \neq \tilde{f}_{12}(x_2(z'))$ ist. Wegen der Stetigkeit von $\tilde{f}_{12}$ gibt es sogar eine ganze Umgebung W von z', so daß für alle $z'' \in W$ gilt: $\tilde{f}_{12}(x_1(z'')) \neq \tilde{f}_{12}(x_2(z''))$. Man kann ein $z'' \in W$ so wählen, daß dazu ein $\tilde{f}_{13} \in \tilde{F}$ existiert mit $\tilde{f}_{13}(x_1(z'')) \neq \tilde{f}_{13}(x_3(z''))$. Auf diese Weise erhalten wir nach endlich vielen Schritten Punkte $x_1, \ldots, x_s$, die über einem Punkt $z_0 \in Z - \tilde{\Phi}(\tilde{N})$ liegen, so daß zu je zwei Punkten $x_{\sigma_1}, x_{\sigma_2} \in \{x_1, \ldots, x_s\}$ eine Abbildung $\tilde{f}_{\sigma_1\sigma_2} \in \tilde{F}$ existiert mit $\tilde{f}_{\sigma_1\sigma_2}(x_{\sigma_1}) \neq \tilde{f}_{\sigma_1\sigma_2}(x_{\sigma_2})$. Die Abbildung $f := (\tilde{f}_{12} \times \tilde{f}_{13} \times \cdots \times \tilde{f}_{s-1,s}) : \tilde{U} \to Y_{f_{12}} \times \times Y_{f_{13}} \times \cdots \times Y_{f_{s-1,s}} =: Y$ bildet dann die Punkte $x_1, \ldots, x_s$ auf s verschiedene Punkte in Y ab. f ist in $\tilde{U}$ stetig, weil die $\tilde{f}_{\sigma_1\sigma_2}$ stetig sind. In dem unverzweigten Teilgebiet $\tilde{U} - \tilde{N}$ ist bereits eine komplexe Struktur erklärt und die Abbildungen aus $\tilde{F}$ sind holomorph bezüglich dieser komplexen Struktur. Daher ist auch f holomorph in $\tilde{U} - \tilde{N}$. $\tilde{\Phi}$ bildet $\tilde{U} - \overset{-1}{\tilde{\Phi}}(\tilde{\Phi}(\tilde{N}))$ eigentlich, lokaltopologisch und holomorph auf $Z - \tilde{\Phi}(\tilde{N})$ ab. Wir können Hilfssatz 3 anwenden: $S(f)$ ist nach Konstruktion von f eine echte Teilmenge von $Z - \tilde{\Phi}(\tilde{N})$.

Der komplexe Raum Y genügt der C-Bedingung[6]). Daher gibt es eine Umgebung A von $f(\tilde{x})$ und endlich viele in A holomorphe Funktionen h_κ, die A separieren. $\overset{-1}{f}(A)$ ist eine Umgebung von $\tilde{x}$. Dann existiert eine Umgebung Z' von $\tilde{\Phi}(\tilde{x})$ mit $\overset{-1}{\tilde{\Phi}}(Z') \cap \tilde{U} \subset \overset{-1}{f}(A)$ (wegen [18], Hilfssatz 1). Nach Hilfssatz 3 ist $S(f)$ eine dünne Teilmenge von Z; daher ist $S(f) \cap Z' \neq Z'$. Also gibt es ein $z_0' \in Z' - S(f) \cap Z'$, und f separiert die Menge $\overset{-1}{\tilde{\Phi}}(z_0') \cap \tilde{U}$. Für $\overset{-1}{\tilde{\Phi}}(Z') \cap \tilde{U} =: \tilde{U}', Z', \lambda(Z') =: E', z_0'$ liegt die gleiche Situation vor wie für $\tilde{U}, Z, E, z_0$. Wir schreiben daher an Stelle von $\tilde{U}', Z', E', z_0'$ wieder $\tilde{U}, Z, E, z_0$. Somit können wir annehmen, daß $f(\tilde{U})$ in A liegt. Eine geeignete Linearkombination h der h_κ separiert die Punkte $f(x_1), \ldots, f(x_s)$. Die Funktion $g := h \circ f$ ist in $\tilde{U}$ stetig und in $\tilde{U} - \tilde{N}$ holomorph. g separiert die Punkte $x_1, \ldots, x_s$.

Wenden wir Hilfssatz 2 auf $\tilde{U} - \overset{-1}{\tilde{\Phi}}(\tilde{\Phi}(\tilde{N})), \tilde{\Phi}, Z - \tilde{\Phi}(\tilde{N})$ an, so folgt, daß g einer Gleichung

$$g^s(x) + c_1(\tilde{\Phi}(x)) \cdot g^{s-1}(x) + \cdots + c_s(\tilde{\Phi}(x)) = 0\,, \quad x \in \tilde{U} - \overset{-1}{\tilde{\Phi}}(\tilde{\Phi}(\tilde{N}))\,,$$

genügt, wobei die Koeffizienten c_σ in $Z - \tilde{\Phi}(\tilde{N})$ holomorph sind. Die Funktion g ist stetig in ganz $\tilde{U}$. Daher sind die c_σ stetig fortsetzbar nach Z und die Gleichung gilt für alle $x \in \tilde{U}$. Die Diskriminante D_g dieser Gleichung ist in Z stetig und in $Z - \tilde{\Phi}(\tilde{N})$ holomorph. In jedem Punkt von $\tilde{\Phi}(\tilde{N})$ verschwindet D_g. Aus dem Satz von Radó-Behnke-Stein [1] folgt, daß D_g in ganz Z holomorph ist. $\tilde{\Phi}(\tilde{N})$ ist im Nullstellengebilde N von D_g enthalten. D_g ist nicht identisch Null, weil g in s übereinanderliegenden Punkten s verschiedene Werte annimmt. Daher ist $N \neq Z$. Damit haben wir eine analytische Menge N mit den gewünschten Eigenschaften konstruiert.

[6]) Vgl. H. Grauert [6].

Die analytisch-verzweigte Überlagerung $(\hat{U}, \lambda, E, (\varphi \circ \mu(N)) \cup M)$ definiert eine komplexe Karte von $\hat{X}$. Die Verträglichkeitsbedingung für zwei auf diese Weise erklärte Karten ist trivialerweise erfüllt, weil sie für die komplexen Karten von B erfüllt ist. Damit ist auf $\hat{X}$ eine komplexe Struktur erklärt. Die Abbildungen aus $\hat{F}$ und $\hat{\Phi}$ sind holomorph bezüglich dieser komplexen Struktur, weil sie in $\hat{X}$ stetig und in $\hat{X}_0$ holomorph sind (s. [16], Satz von RIEMANN). $(\hat{X}, \hat{\Phi}, B)$ ist dann ein Gebiet über B.

2. Es sei (Y, Ψ, B) ein Gebiet über B und $\psi: X \to Y$, $\Phi = \Psi \circ \psi$, eine holomorphe Abbildung, so daß zu jedem $f \in F$ eine holomorphe Abbildung $f': Y \to Y_f$ existiert mit $f = f' \circ \psi$.

Ist Y_0 das Teilgebiet von Y, auf dem Ψ lokaltopologisch abbildet, so gibt es nach Konstruktion von $\hat{X}_0$ eine holomorphe Abbildung $\tau_0: Y_0 \to \hat{X}_0$ mit $\Psi = \hat{\Phi} \circ \tau_0$. Wir beweisen, daß die Abbildung τ_0 zu einer holomorphen Abbildung $\tau: Y \to \hat{X}$, $\Psi = \hat{\Phi} \circ \tau$, fortgesetzt werden kann.

Es sei $y \in Y - Y_0$ und $\mathfrak{U}_y$ der Filter aller offenen zusammenhängenden Umgebungen von y. Dann besteht der Filter $\mathfrak{F} := \mathfrak{U}_y \cap Y_0$ aus offenen zusammenhängenden Mengen und konvergiert gegen y. $\hat{\Phi} \circ \tau_0(\mathfrak{F}) = \Psi(\mathfrak{F})$ konvergiert gegen $\Psi(y) \in B$. Aus Hilfssatz 1 folgt, daß nur zwei Fälle möglich sind:

α) $\tau_0(\mathfrak{F})$ konvergiert gegen einen Punkt von $\hat{X}_0$.

β) $\tau_0(\mathfrak{F})$ besitzt in $\hat{X}_0$ keinen Berührpunkt.

Im Fall α) ist klar, daß τ_0 stetig in den Punkt y fortgesetzt werden kann. Im Fall β) zeigen wir, daß $\tau_0(\mathfrak{F})$ gegen einen Punkt von $\hat{X} - \hat{X}_0$ konvergiert: $\tau_0(\mathfrak{F})$ erfüllt sicher die Forderungen a), b), c) von Definition 1. Ist A eine Umgebung von $\Psi(y)$, so gibt es ein $U \in \mathfrak{U}_y$ mit $\Psi(U) \subset A$. $\tau_0(U \cap Y_0)$ liegt dann in genau einer zusammenhängenden Komponente von $\overset{-1}{\hat{\Phi}}(A)$. Daher ist auch Forderung d) von Definition 1 erfüllt. Also konvergiert $\tau_0(\mathfrak{F})$ gegen einen Randpunkt r von $(\hat{X}_0, \hat{\Phi} \mid \hat{X}_0, B)$. Der Filter $f'(\mathfrak{F})$ konvergiert gegen $f'(y)$; wegen $\hat{f} \circ \tau_0 = f'$ konvergiert auch $\hat{f}(\tau_0(\mathfrak{F}))$, d. h. r gehört zu $\hat{X}_2$. Die Abbildung τ_0 ist damit stetig fortgesetzt zu einer Abbildung $\tau: Y \to \hat{X}_2$. Wir zeigen, daß sogar $\tau(Y) \subset \hat{X}$ gilt. Aus [18], Hilfssatz 3 (s. auch [6], Satz 1) folgt: Es gibt zu jedem Punkt $y \in Y$ eine Umgebung V von y, die durch Ψ eigentlich und nirgends entartet auf eine Umgebung Z von $\Psi(y)$ abgebildet wird. $\tau(V) =: \hat{U}$ ist dann eine Umgebung von $\tau(y)$, die durch $\hat{X}$ eigentlich und nirgends entartet auf Z abgebildet wird. Also gehört $\tau(y)$ zu $\hat{X}$; τ ist die gesuchte Abbildung von Y in $\hat{X}$. Setzt man noch $X = Y$, so ergibt sich die Existenz einer Abbildung $\varphi: X \to \hat{X}$ mit den gewünschten Eigenschaften. Damit ist Satz 1 bewiesen.

Es sei nun (X, Φ, C^n) ein Riemannsches Gebiet und $F := \mathfrak{R}(X)$. Dann folgt aus Satz 1, daß ein maximales Gebiet $\hat{H}(X, \Phi, C^n; \mathfrak{R}(X))$ existiert. $\hat{H}(X, \Phi, C^n; \mathfrak{R}(X))$ besitzt die Eigenschaften der in Definition 2 erklärten Φ-Hülle $\hat{H}(X, \Phi, C^n)$. Es ist klar, daß $\hat{H}(X, \Phi, C^n)$ bis auf spurpunkttreue biholomorphe Abbildungen eindeutig bestimmt ist. Wir erhalten damit als Folgerung von Satz 1:

Satz 2[7]: *Zu jedem Riemannschen Gebiet* (X, Φ, C^n) *existiert genau eine* Φ*-Hülle* $\hat{H}(X, \Phi, C^n)$.

Man definiert: Ein Gebiet (X, Φ, C^n) heißt *holomorph-konvex in einem Randpunkt* r, wenn es eine Umgebung $\overline{U}(r) \subset \overline{X}$ gibt, so daß $\overline{U}(r) \cap X$ holomorph-konvex ist. Verwenden wir den Begriff des *hebbaren Randpunktes*, wie er in [9], Definition 5, erklärt wird, so erhalten wir als notwendige (aber nicht hinreichende) Bedingung für Φ-Hüllen:

Ein Gebiet (X, Φ, C^n), *das mit seiner* Φ*-Hülle übereinstimmt, ist in jedem hebbaren Randpunkt holomorph-konvex. Es gibt eine höchstens* $(n-1)$*-dimensionale analytische Menge* M *in* X, *so daß jeder Randpunkt, in dem* (X, Φ, C^n) *nicht holomorph-konvex ist, Häufungspunkt von* M *ist.*

Beweis: Das Teilgebiet X_0 von X, auf dem Φ lokaltopologisch abbildet, ist nach Konstruktion der Φ-Hülle ein unverzweigtes holomorph-konvexes Riemannsches Gebiet. Zu jedem hebbaren Randpunkt $r \in RdX$ gibt es eine schlichte Umgebung $\overline{U}(r) \subset \overline{X}$. Daher ist $\overline{U} \cap X \subset X_0$. r ist also auch Randpunkt von $(X_0, \Phi \mid X_0, C^n)$, somit ist (X, Φ, C^n) in einer Umgebung von r holomorph-konvex. Wenn (X, Φ, C^n) in einem Randpunkt r' nicht holomorph-konvex ist, dann gilt notwendig $\overline{U} \cap (X - X_0) \neq \emptyset$ für jede Umgebung $\overline{U}$ von r'. Also ist r' Häufungspunkt der Verzweigungsmenge $M := X - X_0$.

Wir wollen noch ein *Beispiel* angeben[8]:

Es sei $X := \{(z_1, z_2, z_3) : |z_1|^2 + |z_2|^2 > 1\}$ und Φ die Abbildung $w_1 = (z_1 + z_2) \cdot z_3$, $w_2 = z_1 \cdot (1 + z_3)$, $w_3 = z_2 \cdot (1 - z_3)$ von X in C^3_w. $\Phi : X \to C^3_w$ ist nirgends entartet, daher ist (X, Φ, C^3_w) ein (verzweigtes) Gebiet über C^3_w. Es ist $\hat{H}(X, \Phi, C^3_w) = \{(z_1, z_2, z_3) : (z_1, z_2) \neq (0, 0)\}$. Bezeichnet nämlich ι die Injektion von X in C^3_z, so ist $\hat{H}(X, \iota, C^3_z) = C^3_z$. Die Abbildung $\Phi : C^3_z \to C^3_w$ ist in der Menge $\{(z_1, z_2, z_3) : z_1 = z_2 = 0\}$ entartet. Setzen wir $Y := \{(z_1, z_2, z_3) : (z_1, z_2) \neq (0, 0)\}$, so stimmt das Gebiet (Y, Φ, C^3_w) mit seiner Φ-Hülle $\hat{H}(Y, \Phi, C^3_w)$ überein, dagegen ist $\hat{H}(Y, \iota, C^3_z) = C^3_z \neq Y$.

Dieses Beispiel zeigt, daß die Φ-Hülle von der speziellen Projektionsabbildung Φ abhängt. Daher werden wir eine allgemeinere Holomorphiehülle einführen, die nur von dem Raum X abhängig ist.

§ 2. Die Existenz der K-Hülle

Ein komplexer Raum X heißt K-vollständig, wenn es zu jedem Punkt $x \in X$ endlich viele in X holomorphe Funktionen $f_1, \ldots, f_n$ gibt, so daß die durch $(f_1, \ldots, f_n) : X \to C^n$ gegebene Abbildung in einer Umgebung von x nirgends entartet ist.

Wir wollen zu jedem K-vollständigen komplexen Raum eine Holomorphiehülle erklären:

Definition 3: *Ein komplexer Raum* $H(X)$ *heißt eine* K*-Hülle des komplexen Raumes* X, *wenn gilt:*

[7] Vgl. [1].
[8] Vgl. [10].

Es existiert eine nirgends entartete holomorphe Abbildung $\alpha: X \to H(X)$, so daß der induzierte Homomorphismus $^*\alpha: \mathfrak{R}(H(X)) \to \mathfrak{R}(X)$ ein Isomorphismus von $\mathfrak{R}(H(X))$ auf $\mathfrak{R}(X)$ ist.

b) *Zu jedem K-vollständigen komplexen Raum Y, zu dem es eine nirgends entartete holomorphe Abbildung $\beta: X \to Y$ gibt, die einen Isomorphismus $^*\beta$ von $\mathfrak{R}(Y)$ auf $\mathfrak{R}(X)$ induziert, existiert eine nirgends entartete holomorphe Abbildung $\gamma: Y \to H(X)$ mit $\alpha = \gamma \circ \beta$.*

Wir beweisen folgenden Satz:

Satz 3: *Zu jedem K-vollständigen komplexen Raum X existiert eine K-Hülle $H(X)$. $H(X)$ ist bis auf analytische Isomorphie eindeutig bestimmt. $H(X)$ ist K-vollständig.*

Dem Beweis von Satz 3 stellen wir zwei Hilfssätze voran.

Es seien $\varphi: X \to C^n$ und $\psi: X \to C^n$ zwei nirgends entartete holomorphe Abbildungen von X in einen C^n. Wir bilden die φ-Hülle $\hat{H}(X, \varphi, C^n)$ und die ψ-Hülle $\hat{H}(X, \psi, C^n)$. Die Abbildung φ ist durch ein n-tupel holomorpher Funktionen gegeben. φ ist daher nach $\hat{H}(X, \psi, C^n)$ fortsetzbar zu einer holomorphen Abbildung $\tilde{\varphi}: \hat{H}(X, \psi, C^n) \to C^n$. Es sei $M(\psi, \varphi)$ die Entartungsmenge von $\tilde{\varphi}$ in $\hat{H}(X, \psi, C^n)$. Dann ist $G(\psi, \varphi) := \hat{H}(X, \psi, C^n) - M(\psi, \varphi)$ ein offenes Teilgebiet von $\hat{H}(X, \psi, C^n)$[9]). Entsprechend sei $G(\varphi, \psi)$ das Teilgebiet von $\hat{H}(X, \varphi, C^n)$, das durch $\tilde{\psi}$ nirgends entartet abgebildet wird. Dann gilt:

Hilfssatz 4: *Es gibt eine und nur eine biholomorphe Abbildung τ von $G(\varphi, \psi)$ auf $G(\psi, \varphi)$, die mit den vorgegebenen Abbildungen kommutiert.*

Hilfssatz 5: *Ist x_0 ein Punkt von $M(\varphi, \psi)$ und $\mathfrak{U}_{x_0}$ der Umgebungsfilter von x_0 in $\hat{H}(X, \varphi, C^n)$, so besitzt der Filter $\tau(\mathfrak{U}_{x_0} \cap G(\varphi, \psi))$ keinen Berührpunkt in $\hat{H}(X, \psi, C^n)$.*

Beweis von Hilfssatz 4: Es sei φ_1 die Abbildung von X in $\hat{H}(X, \varphi, C^n)$, φ_2 die Projektionsabbildung von $\hat{H}(X, \varphi, C^n)$ in den C^n, $\varphi = \varphi_2 \circ \varphi_1$. Entsprechend seien $\psi_1: X \to H(X, \psi, C^n)$ und $\psi_2: \hat{H}(X, \psi, C^n) \to C^n$, $\psi = \psi_2 \circ \psi_1$, erklärt. $(G(\varphi, \psi), \tilde{\psi}, C^n)$ ist ein Gebiet, in das alle Funktionen aus $\mathfrak{R}(X)$ fortsetzbar sind. Nach Definition der ψ-Hülle gibt es genau eine holomorphe Abbildung τ von $G(\varphi, \psi)$ in $\hat{H}(X, \psi, C^n)$ mit $\tau \circ \varphi_1 = \psi_1$ und $\psi_2 \circ \tau = \tilde{\psi}$. Ebenso existiert eine holomorphe Abbildung σ von $G(\psi, \varphi)$ in $\hat{H}(X, \varphi, C^n)$ mit $\sigma \circ \psi_1 = \varphi_1$ und $\varphi_2 \circ \sigma = \tilde{\varphi}$. Es ist sogar $\tau(G(\varphi, \psi)) \subset G(\psi, \varphi)$, weil φ_2 in $G(\varphi, \psi)$, also auch $\tilde{\varphi} = \varphi_2 \circ \tau$ in $\tau(G(\varphi, \psi))$ nirgends entartet ist. Daher ist τ eine Abbildung von $G(\varphi, \psi)$ in $G(\psi, \varphi)$; ebenso gilt $\sigma: G(\psi, \varphi) \to G(\varphi, \psi)$. Es gilt $(\tau \circ \sigma) \circ \psi_1 = \psi_1$ und $(\sigma \circ \tau) \circ \varphi_1 = \varphi_1$. Daraus folgt $\tau = \sigma^{-1}$; τ ist also eine umkehrbar holomorphe Abbildung von $G(\varphi, \psi)$ auf $G(\psi, \varphi)$.

Beweis zu Hilfssatz 5: Die Abbildung τ ist eine topologische Abbildung von $G(\varphi, \psi)$ auf $G(\psi, \varphi)$. Daraus folgt, daß $\tau(\mathfrak{U}_{x_0} \cap G(\varphi, \psi))$ sicher in $G(\psi, \varphi)$ keinen Berührpunkt besitzt. Wir haben noch zu zeigen, daß $\tau(\mathfrak{U}_{x_0} \cap G(\varphi, \psi))$ auch keinen Punkt von $M(\psi, \varphi)$ als Berührpunkt hat. Wir dürfen annehmen, daß $\mathfrak{U}_{x_0}$ der Filter aller offenen zusammenhängenden Umgebungen von x_0 in $\hat{H}(X, \varphi, C^n)$ ist. Dann besteht der Filter $\mathfrak{U}_{x_0} \cap G(\varphi, \psi)$ aus offenen zusammen-

[9]) Vgl. [16], Satz 18.

hängenden Mengen und das gleiche gilt für den Filter $\mathfrak{F} := \tau(\mathfrak{U}_{x_0} \cap G(\varphi, \psi))$, weil $M(\varphi, \psi)$ eine nirgends zerlegende Teilmenge von $\hat{H}(X, \varphi, C^n)$[10]) und τ eine topologische Abbildung ist. Die Abbildung $\tilde{\psi}$ ist in ganz $\hat{H}(X, \varphi, C^n)$ stetig; wegen $\tilde{\psi} = \psi_2 \circ \tau$ konvergiert der Filter $\psi_2(\mathfrak{F})$ gegen $\tilde{\psi}(x_0)$. Nehmen wir an, $\mathfrak{F}$ besitze in $M(\psi, \varphi)$ einen Berührpunkt. Dann folgt aus Hilfssatz 1, daß $\mathfrak{F}$ gegen einen Punkt $y_0 \in M(\psi, \varphi)$ konvergiert. Wählen wir eine kompakte Umgebung K von y_0, so gibt es eine Umgebung U von x_0, so daß $\tau(U \cap G(\varphi, \psi))$ in K liegt. Ist dann $x \in U \cap M(\varphi, \psi)$ und $\mathfrak{U}_x$ der Filter aller offenen zusammenhängenden Umgebungen von x, so besitzt $\tau(\mathfrak{U}_x \cap G(\varphi, \psi))$ in $K \subset \hat{H}(X, \psi, C^n)$ einen Berührpunkt. Nochmalige Anwendung von Hilfssatz 1 ergibt, daß τ in eine Umgebung $U(x_0)$ stetig fortsetzbar ist. Ebenso beweist man, daß die Abbildung τ^{-1} in eine Umgebung von y_0 stetig fortsetzbar ist. (Man ersetze τ, x_0, ψ durch τ^{-1}, y_0, φ und wende darauf die gleiche Schlußweise an.) Also bildet τ eine Umgebung von x_0 eineindeutig auf eine Umgebung von y_0 ab. Die Abbildung φ_2 ist in x_0 nicht entartet. Aus $\varphi_2 = \tilde{\varphi} \circ \tau$ folgt, daß dann auch $\tilde{\varphi}$ in y_0 nicht entartet ist. y_0 ist aber ein Punkt von $M(\psi, \varphi)$, daher ist y_0 eine Entartungsstelle von $\tilde{\varphi}$. Damit ist ein Widerspruch hergestellt und Hilfssatz 5 bewiesen.

Nun zum *Beweis von Satz 3:*

Wir bilden $R := \bigcup_{\varphi} \hat{H}(X, \varphi, C^n)$, wobei φ alle nirgends entarteten holomorphen Abbildungen von X in den C^n durchlaufe. Dann erklären wir in R eine Äquivalenzrelation Z:

Sind x und y zwei Punkte aus R, so gibt es ein φ und ein ψ, so daß gilt: $x \in \hat{H}(X, \varphi, C^n)$, $y \in \hat{H}(X, \psi, C^n)$. Wir bilden die Gebiete $G(\varphi, \psi)$ und $G(\psi, \varphi)$. Dazu existiert nach Hilfssatz 4 in eindeutiger Weise die Abbildung $\tau : G(\varphi, \psi) \to G(\psi, \varphi)$. Wir nennen die Punkte x und y Z-äquivalent ($x \sim y$), wenn $x \in G(\varphi, \psi)$ und $y \in G(\psi, \varphi)$ ist und gilt: $y = \tau(x)$.

Es ist klar, daß Z wirklich eine Äquivalenzrelation ist.

Wir zeigen nun, daß R/Z ein Hausdorffscher Raum ist. Dies ist dann der Fall, wenn gilt[11]):

1) *Z ist eine offene Zerlegung.* — Es sei A eine offene Teilmenge von R. Dann ist die saturierte Hülle von A die Vereinigung von Mengen $\tau(A \cap G(\varphi, \psi))$. Jedes $G(\varphi, \psi)$ ist offen in R, also ist $\tau(A \cap G(\varphi, \psi))$ offen und damit auch die saturierte Hülle von A.

2) *Die durch Z bestimmte Menge $C := \{(x, y) \in R \times R : x \sim y\}$ ist abgeschlossen in $R \times R$.* — Dies ist offenbar dann der Fall, wenn gilt: Sind x und y nicht Z-äquivalent, so gibt es Umgebungen $U(x)$ und $V(y)$ in R, so daß kein Punkt von U zu einem Punkt von V Z-äquivalent ist. Wir nehmen an, dies sei nicht erfüllt. Dann gibt es zwei Punkte $x_0, y_0 \in R$, die nicht Z-äquivalent sind, und dazu Punkte $x_i, y_i \in R$ mit $\lim x_i = x_0$, $\lim y_i = y_0$ und $x_i \sim y_i$ (i durchlaufe eine gerichtete Indexmenge). Das bedeutet: Es ist $x_0 \in \hat{H}(X, \varphi, C^n)$ und $y_0 \in \hat{H}(X, \psi, C^n)$, $\tau : G(\varphi, \psi) \to G(\psi, \varphi)$, $\tau(x_i) = y_i$. (Wir können an-

---

[10]) Vgl. [16], Satz 18.

[11]) Vgl. [3], § 7, 18.

nehmen, daß für alle i gilt: $x_i \in G(\varphi, \psi)$, $y_i \in G(\psi, \varphi)$.) Ist nun $x_0 \in G(\varphi, \psi)$, dann ist notwendig $y_0 = \lim y_i = \lim \tau(x_i) = \tau(x_0)$; also $x_0 \sim y_0$, im Gegensatz zur Voraussetzung, daß x_0 und y_0 nicht Z-äquivalent sein sollen. Ist dagegen $x_0 \in M(\varphi, \psi)$, so besitzt nach Hilfssatz 5 $\tau(x_i) = y_i$ keinen Häufungspunkt in $\mathfrak{H}(X, \psi, C^n)$. Dies widerspricht der Voraussetzung $\lim y_i = y_0$.

Damit ist gezeigt, daß R/Z ein Hausdorffscher Raum ist. Die Abbildungen τ, durch die die Äquivalenzrelation Z erklärt ist, sind biholomorph. Daher trägt R/Z in natürlicher Weise eine komplexe Struktur. Wir setzen $R/Z =: H(X)$ und zeigen, daß R/Z die Eigenschaften einer K-Hülle besitzt:

a) X ist K-vollständig; nach einem Satz von H. GRAUERT ([6], Satz A) gibt es eine nirgends entartete holomorphe Abbildung $\varphi : X \to C^n$. Es gibt nach Definition der φ-Hülle eine Abbildung von X in $\mathfrak{H}(X, \varphi, C^n)$, außerdem eine Abbildung von $\mathfrak{H}(X, \varphi, C^n)$ in R und die Abbildung von R in R/Z. Dadurch ist eine Abbildung α von X in R/Z gegeben, die die verlangte Eigenschaft besitzt.

b) Ist Y ein K-vollständiger komplexer Raum mit der Eigenschaft b) von Definition 3, so existiert zu Y eine nirgends entartete holomorphe Abbildung $\psi : Y \to C^n$ ([6], Satz A). Dann ist die Abbildung $\psi \circ \beta : X \to C^n$ ebenfalls nirgends entartet. Es gibt eine Abbildung von Y in $\mathfrak{H}(X, \psi \circ \beta, C^n)$. Die durch $Y \to \mathfrak{H}(X, \psi \circ \beta, C^n) \to R \to R/Z$ gegebene holomorphe Abbildung $\gamma : Y \to R/Z$ erfüllt die Forderung b) von Definition 3.

Aus der Konstruktion von $H(X)$ ersieht man leicht, daß $H(X)$ K-vollständig ist. Aus Definition 3 folgt, daß zwei K-Hüllen analytisch äquivalent sind. Damit ist Satz 3 bewiesen.

Es ergibt sich unmittelbar, daß die K-Hülle invariant gegenüber biholomorphen Abbildungen ist:

Ist $\lambda : X_1 \to X_2$ eine biholomorphe Abbildung des K-vollständigen komplexen Raumes X_1 auf den K-vollständigen komplexen Raum X_2, so gibt es eine biholomorphe Abbildung $\tilde{\lambda} : H(X_1) \to H(X_2)$ von $H(X_1)$ auf $H(X_2)$, so daß gilt $\alpha_2 \circ \lambda = \tilde{\lambda} \circ \alpha_1$. ($\alpha_i$ sei die Abbildung von X_i in $H(X_i)$, $i = 1, 2$.)

§ 3. Φ-Hülle und K-Hülle

Es sei (X, Φ, C^n) ein Riemannsches Gebiet. Dann ist X ein K-vollständiger komplexer Raum. Wir können daher die beiden Hüllen $\mathfrak{H}(X, \Phi, C^n)$ und $H(X)$ bilden. Es gilt:

Satz 4: *$\mathfrak{H}(X, \Phi, C^n)$ ist ein Teilgebiet von $H(X)$, und zwar ist $H(X) - \mathfrak{H}(X, \Phi, C^n)$ genau die Entartungsmenge der nach $H(X)$ fortgesetzten Abbildung Φ.*

Beweis: Die Abbildung $\Phi : X \to C^n$ ist fortsetzbar zu einer holomorphen Abbildung $\hat{\Phi} : H(X) \to C^n$. Das Teilgebiet von $H(X)$, das durch $\hat{\Phi}$ nirgends entartet abgebildet wird, besitzt offenbar die Eigenschaften der Φ-Hülle; es stimmt also mit $\mathfrak{H}(X, \Phi, C^n)$ überein. Somit ist $H(X) - \mathfrak{H}(X, \Phi, C^n)$ die Entartungsmenge von $\hat{\Phi}$.

Aus einem Satz von R. Remmert ([16], Satz 18) folgt dann:

Satz 5: *$H(X) - \hat{H}(X, \Phi, C^n)$ ist eine höchstens $(n-1)$-dimensionale analytische Menge in $H(X)$, die keine isolierten Punkte enthält.*

H. Grauert und R. Remmert haben an Beispielen gezeigt, daß eine Φ-Hülle $\hat{H}(X, \Phi, C^n)$ nicht holomorph-konvex zu sein braucht [9]. Es gilt:

Satz 6: *Ist $\hat{H}(X, \Phi, C^n) \neq H(X)$, so ist $\hat{H}(X, \Phi, C^n)$ nicht holomorph-konvex.*

Beweis: Es sei $x_0 \in H(X) - \hat{H}(X, \Phi, C^n)$ und $x_\nu \in \hat{H}(X, \Phi, C^n)$ eine Punktfolge mit $\lim x_\nu = x_0$. Dann besitzt die Folge x_ν in $\hat{H}(X, \Phi, C^n)$ keinen Häufungspunkt. Jede in $\hat{H}(X, \Phi, C^n)$ holomorphe Funktion f ist auch in $H(X)$ holomorph und daher in einer Umgebung von x_0 beschränkt. Daher ist $|f(x_\nu)|$ beschränkt und somit $\hat{H}(X, \Phi, C^n)$ nicht holomorph-konvex.

Wie das Beispiel am Ende von § 1 zeigt, gibt es K-vollständige komplexe Räume Y, die bei einer bestimmten Konkretisierung als Riemannsches Gebiet (Y, Φ, C^n) mit ihrer Φ-Hülle $\hat{H}(Y, \Phi, C^n)$ übereinstimmen, bei einer anderen Konkretisierung durch eine Abbildung ι aber von $\hat{H}(Y, \iota, C^n)$ verschieden sind. Die K-Hüllen sind dadurch charakterisiert, daß diese Erscheinung nicht auftritt.

Satz 7: *Für einen K-vollständigen komplexen Raum X ist dann und nur dann $X = H(X)$, wenn für jede nirgends entartete holomorphe Abbildung $\Phi: X \to C^n$ gilt: $X = \hat{H}(X, \Phi, C^n)$.*

Der Beweis ist trivial. Aus der Konstruktion von $\hat{H}(X, \Phi, C^n)$ und Satz 7 folgt:

Ist $X = H(X)$ und $\Phi: X \to C^n$ eine beliebige nirgends entartete holomorphe Abbildung, so ist das Teilgebiet von X, auf dem Φ lokaltopologisch ist, ein holomorph-vollständiger komplexer Raum.

Wir zeigen jetzt, daß beide Hüllenbegriffe, die Φ-Hülle und die K-Hülle, Verallgemeinerungen der von P. Thullen konstruierten unverzweigten Hülle eines unverzweigten Riemannschen Gebietes sind.

Satz 8: *Ist (X, Φ, C^n) ein unverzweigtes Riemannsches Gebiet, dann gilt immer $\hat{H}(X, \Phi, C^n) = H(X)$. Beide Hüllen stimmen mit der unverzweigten Holomorphiehülle (im Sinne von* P. Thullen *[19]) überein.*

Beweis: Nach Definition 3, b) gibt es eine Abbildung γ der unverzweigten Holomorphiehülle $H_0(X)$ in $H(X)$. γ ist injektiv, weil $H_0(X)$ holomorph-separabel ist. Außerdem ist γ eigentlich: Ist K eine kompakte Teilmenge von $H(X)$, so ist jede in $H(X)$ holomorphe Funktion auf K beschränkt. Daher ist jede in $H_0(X)$ holomorphe Funktion auf $\overset{-1}{\gamma}(K)$ beschränkt. $H_0(X)$ ist holomorph-konvex. Aus einem Satz von H. Cartan ([4], IX, théorème 1; s. auch R. Remmert [14]) folgt, daß $\overset{-1}{\gamma}(K)$ kompakt ist. Also ist γ eigentlich und daher surjektiv. Damit ist gezeigt, daß γ biholomorph ist, d. h. $H_0(X) = H(X)$. Eine lokaltopologische Abbildung ist insbesondere nirgends entartet. Dann folgt aus Satz 4, daß $\hat{H}(X, \Phi, C^n) = H(X)$ ist.

Wir wollen noch ein Beispiel angeben:

Beispiel: Es sei (G, φ, C^n) ein unverzweigtes Riemannsches Gebiet; $\varphi: G \to C^n$ wird also als lokaltopologisch vorausgesetzt. Weiter sei $\Phi: G \to C^n$ eine nirgends entartete holomorphe Abbildung. (G, Φ, C^n) ist dann ein Riemannsches Gebiet. Nach Satz 8 gilt $\hat{H}(G) = \hat{H}(G, \varphi, C^n)$. Die K-Hülle $H(G)$ ist also holomorph-vollständig. Aus einem Satz von H. GRAUERT und R. REMMERT ([9], Satz 7) folgt, daß $H(G) - \hat{H}(G, \Phi, C^n)$ eine höchstens $(n-2)$-dimensionale analytische Menge in $H(G)$ ist. Jeder nicht-nulldimensionalen Niveaumenge der in die K-Hülle fortgesetzten Abbildung Φ entspricht ein Randpunkt von $\hat{H}(G, \Phi, C^n)$, in dem die Φ-Hülle $\hat{H}(G, \Phi, C^n)$ nicht holomorph-konvex ist. Alle Randpunkte, in denen $\hat{H}(G, \Phi, C^n)$ nicht holomorph-konvex ist, entstehen auf diese Weise. Für die Dimension $n = 2$ ergibt sich in unserem Beispiel: Für jede nirgends entartete holomorphe Abbildung Φ ist $\hat{H}(G, \Phi, C^2) = H(G)$. In diesem Fall ist also jede Φ-Hülle $\hat{H}(G, \Phi, C^2)$ holomorph-vollständig.

Literatur

[1] BEHNKE, H., u. K. STEIN: Modifikation komplexer Mannigfaltigkeiten und Riemannscher Gebiete. Math. Ann. **124**, 1—16 (1951). — [2] BEHNKE, H., u. P. THULLEN: Theorie der Funktionen mehrerer komplexen Veränderlichen. Ergebn. Math. **3**, 3 (1934). — [3] BOURBAKI, N.: Topologie générale. Fascicule de résultats. Paris: Hermann & Cie. 1951. — [4] CARTAN, H.: Séminaire. Paris 1951/52. — [5] CARTAN, H., u. P. THULLEN: Zur Theorie der Singularitäten der Funktionen mehrerer komplexer Veränderlichen. Regularitäts- und Konvergenzbereiche. Math. Ann. **106**, 617—647 (1932). — [6] GRAUERT, H.: Charakterisierung der holomorph vollständigen komplexen Räume. Math. Ann. **129** 233—259 (1955). — [7] GRAUERT, H.: Charakterisierung der Holomorphiegebiete durch die vollständige Kählersche Metrik. Math. Ann. **131**, 38—75 (1956). — [8] GRAUERT, H., u. R. REMMERT: Zur Theorie der Modifikationen. I. Stetige und eigentliche Modifikationen komplexer Räume. Math. Ann. **129**, 274—296 (1955). — [9] GRAUERT, H., u. R. REMMERT: Konvexität in der komplexen Analysis. Nicht holomorph-konvexe Holomorphiegebiete und Anwendungen auf die Abbildungstheorie. Comm. Math. Helv. **31**, 152—183 (1956). — [10] GRAUERT, H., u. R. REMMERT: Singularitäten komplexer Mannigfaltigkeiten und Riemannsche Gebiete. Math. Z. **67**, 103—128 (1957). — [11] GRAUERT, H., u. R. REMMERT: Komplexe Räume. Math. Ann. **136**, 245—318 (1958). — [12] IWAHASHI, R.: Domains spread on a complex space. J. Math. Soc. Japan **9**, 4 (October 1957). — [13] REMMERT, R.: Projektionen analytischer Mengen. Math. Ann. **130**, 410—441 (1956). — [14] REMMERT, R.: Sur des espaces analytiques holomorphiquement séparables et holomorphiquement convexes. C. R. Acad. Sci. (Paris) **243**, 118—121 (1956). — [15] REMMERT, R.: Habilitationsschrift. Münster 1956. — [16] REMMERT, R.: Holomorphe und meromorphe Abbildungen komplexer Räume. Math. Ann. **133**, 328—370 (1957). — [17] SCHEJA, G.: Verzweigte Holomorphiehüllen. S.-B. bayr. Akad. Wiss. 1958. — [18] STEIN, K.: Analytische Zerlegungen komplexer Räume. Math. Ann. **132**, 63—93 (1956). — [19] THULLEN, P.: Zur Theorie der Funktionen zweier komplexer Veränderlichen. Die Regularitätshüllen. Math. Ann. **106**, 64—72 (1932).

(Eingegangen am 30. April 1959)

GOLDBERG, S.
Math. Annalen 138, 329—331 (1959)

Some Properties of the Space of Compact Operators on a Hilbert Space

By

SEYMOUR GOLDBERG in Jerusalem, Israel

Introduction

In [4], SCHATTEN proves that the space of compact operators, $\mathscr{K}$, mapping a Hilbert space $\mathscr{H}$ into itself, is not equivalent to the conjugate space X' of any Banach space X. It is shown in this paper that if $\mathscr{H}$ is separable, then neither $\mathscr{K}$ nor $\mathscr{K}_N$, the subspace of compact normal operators, are even isomorphic to X'. It is also shown that the space of bounded operators on $\mathscr{H}$ (separable) cannot be projected onto $\mathscr{K}$. Finally, some properties of $\mathscr{K}$ and its conjugate space $\mathscr{K}'$ are listed, when $\mathscr{H}$ is not necessarily separable.

Lemma. If $\mathscr{H}$ is a separable Hilbert space, then $\mathscr{K}$ is also separable.

Proof: As in [5], if φ and ψ are two elements of $\mathscr{H}$, let $\varphi \otimes \overline{\psi}$ be the operator defined by $(\varphi \otimes \overline{\psi}) f = (f, \psi)\, \varphi$, $f \in \mathscr{H}$. For each operator A of finite rank n, vectors $\eta_1, \eta_2, \ldots, \eta_n, \psi_1, \psi_2, \ldots, \psi_n$ can be selected so that

$$A = \sum_1^n \eta_i \otimes \overline{\psi}_i \,. \tag{1}$$

Since the space $\mathscr{K}_1$, of operators of finite rank is dense in $\mathscr{K}$, we need only consider $\mathscr{K}_1$. Let $\varphi_1, \varphi_2, \ldots$ be a complete orthonormal set in $\mathscr{H}$. Then all operators of the form

$$\sum_{i=1}^m \sum_{j=1}^n \alpha_{ij} \varphi_i \otimes \overline{\varphi}_j \,, \tag{2}$$

where the α_{ij} are "rational" (i.e. real and imaginary parts of α_{ij} are rational), is denumerable. For a non zero operator $\eta \otimes \overline{\psi}$ and any $\varepsilon > 0$, an element $\psi_\varepsilon = \sum_1^N \beta_j \varphi_j$, β_j rational, can be chosen so that $\|\psi - \psi_\varepsilon\| < \frac{\varepsilon}{2\|\eta\|}$ and an element $\eta_\varepsilon = \sum_1^M \gamma_i \varphi_i$, γ_i rational, so that $\|\eta - \eta_\varepsilon\| < \frac{\varepsilon}{2\|\psi_\varepsilon\|}$. Then for any $f \in \mathscr{H}$, $\|f\| \leq 1$, it follows that

$$\|(f, \psi)\, \eta - (f, \psi_\varepsilon)\, \eta_\varepsilon\| \leq |(f, \psi - \psi_\varepsilon)|\, \|\eta\| + |(f, \psi_\varepsilon)|\, \|\eta - \eta_\varepsilon\| < \varepsilon .$$

Hence $\left\| \eta \otimes \overline{\psi} - \sum_{i=1}^M \sum_{j=1}^N \gamma_i \overline{\beta}_j \varphi_i \otimes \overline{\varphi}_j \right\| < \varepsilon$. Clearly, we may now approximate A in (1) by an element in (2).

Theorem. If $\mathscr{H}$ is a separable Hilbert space, then neither $\mathscr{K}$ nor $\mathscr{K}_N$ is isomorphic to X'.

Proof. Let $\varphi_1, \varphi_2, \ldots$ be a complete orthonormal set in $\mathscr{H}$. For $\lambda = \{\lambda_i\}$ in c_0, the space of sequences converging to zero, we define $h(\lambda) = T_\lambda$, where $T_\lambda\left(\sum_1^\infty \alpha_i \varphi_i\right) = \sum_1^\infty \lambda_i \alpha_i \varphi_i$. By ([6] p. 359), T_λ is normal and compact; i.e. $T_\lambda \in K_N$. The operator h is clearly linear. Moreover, it is also an isometry, for if $x = \sum_1^\infty \alpha_i \varphi_i$, $\|x\| \leq 1$, then $\|T_\lambda x\| = \left\|\sum_1^\infty \lambda_i \alpha_i \varphi_i\right\| \leq \sup_j |\lambda_j| \, \|x\| \leq \|\lambda\|$. Suppose $|\lambda_k| = \sup |\lambda_j| = \|\lambda\|$. Then $\|T_\lambda \varphi_k\| = \|\lambda_k \varphi_k\| = \|\lambda\|$. Hence $\|h(\lambda)\| = \|T_\lambda\| = \|\lambda\|$, so that h defines an equivalence between c_0 and a subspace of $\mathscr{K}_N$. If $\mathscr{K}_N(\mathscr{K})$ were isomorphic to X', then X' would obviously contain a subspace isomorphic to c_0. But by [1] th. 4, X' and hence $\mathscr{K}_N(\mathscr{K})$ must contain a subspace isomorphic to m, the space of bounded sequences. It is well known however, that m is not separable, which implies that $\mathscr{K}_N(\mathscr{K})$ cannot be separable. Thus we have a contradiction of the lemma.

Corollary 1. If $\mathscr{H}$ is separable, then the conjugate spaces $\mathscr{K}'$ and $\mathscr{K}'_N$ do not contain a minimal total closed subspace.

Proof: This is a direct result of the above theorem and [3] th. 1.

Remarks 1. The author conjectures that the theorem is true even when $\mathscr{H}$ is not separable.

2. If $\mathscr{H}$ is not separable, then c_0 can still be imbedded into $\mathscr{K}_N$ as follows:

Choose an infinite orthonormal set $\varphi_1, \varphi_2, \ldots$ in $\mathscr{H}$. Let Φ be the closed linear manifold generated by the φ_i, and let $\Phi^\perp$ be the orthogonal complement of Φ. We define T_λ as in the theorem and extend T_λ to $\tilde{T}_\lambda$ defined on all of $\mathscr{H}$ by setting $\tilde{T}_\lambda(x + y) = T_\lambda x$, $x \in \Phi$, $y \in \Phi^\perp$. It is easy to see that $\tilde{T}_\lambda \in \mathscr{K}_N$. The above theorem can now be generalized as in the following corollary:

Corollary 2. If $\mathscr{M}$ is a linear manifold of operators defined on $\mathscr{H}$ (not necessarily separable) and $\mathscr{M}$ contains the T_λ but no subspace isomorphic to m, then $\mathscr{M}$ cannot be isomorphic to X'.

The fact that c_0 can be embedded in $\mathscr{K}_N$ enables us to freely make use of some theorems in [1] and [2].

Corollary 3. $\mathscr{K}_N$, $\mathscr{K}$ and the space $\mathscr{B}$ of bounded operators on $\mathscr{H}$ can be projected onto a subspace isomorphic to l. ([2] cor. 2.)

Corollary 4. If $\mathscr{H}$ is separable, then $\mathscr{K}_N$ and $\mathscr{K}$ can, in addition, be projected onto a subspace isomorphic to c_0. ([1] c. 6.)

Corollary 5. The conjugate spaces $\mathscr{K}'_N$, $\mathscr{K}'$ and $\mathscr{B}'$ can be projected onto a subspace isomorphic to l.

Proof: Since h maps c_0 into $\mathscr{K}_N \subset \mathscr{K} \subset \mathscr{B}$, the conjugate operator h' can be used to map $\mathscr{B}'$, $\mathscr{K}'$, $\mathscr{K}'_K$ onto l. The corollary now follows from [2] th. 1 (iii) and (iv).

Corollary 6. The space l is equivalent to a subspace of $\mathscr{K}'$.

Proof: Let $\mathscr{I}$ be the set of all operators A on $\mathscr{H}$ for which

$$\sum_i ((A^*A)^{1/2}\, \eta_i, \eta_i) < \infty \tag{1}$$

for a complete orthonormal system. This sum is independent of the complete orthonormal system chosen. $\mathscr{I}$ may be interpreted as $\mathscr{K}'$ when the sum in (1) stands for the norm of A in $\mathscr{I}$. (For a complete discussion of $\mathscr{I}$, see [5], pp. 110—116.)

We choose an infinite orthonormal system $\varphi_1, \varphi_2, \ldots$ (the system is not necessarily complete). As in Remark 2, operators T_i are defined by the relations $T_i\varphi_i = 1$, $T_i\varphi_j = 0$, $i \neq j$ and are extended to all of $\mathscr{H}$. For each $\alpha = \{\alpha_i\}$ in l, $A_\alpha = \sum_1^\infty \alpha_i T_i$ is well defined. Since the T_i are self adjoint and $T_i T_j = 0$ $i \neq j$, it follows that

$$(A_\alpha^* A_\alpha)^{1/2} = \sum_l |\alpha_i|\, T_i .$$

Hence

$$\sum_l^\infty ((A_\alpha^* A_\alpha)^{1/2}\, \varphi_j, \varphi_j) = \sum_l^\infty |\alpha_i| = \|\alpha\| .$$

The mapping h defined by $h(\alpha) = A_\alpha$ is thus a linear isometry from l into $\mathscr{I} = \mathscr{K}'$.

Corollary 7. $\mathscr{B}$ contains a subspace, $\mathscr{B}_1$, isomorphic to c_0, but $\mathscr{B}$ cannot be projected onto $\mathscr{B}_1$.

Proof: By [2] cor. 4 and the fact that $\mathscr{B}$ can be identified with $\mathscr{K}''$, ([5] th. 5.15).

Remark: If $\mathscr{H}$ is separable, then corollaries 4 and 7 show that $\mathscr{K}$ can be projected onto a subspace isomorphic to c_0 but $\mathscr{B}$ cannot.

Theorem. If $\mathscr{H}$ is separable, then the space $\mathscr{B}$ of bounded linear operators on $\mathscr{H}$ cannot be projected onto the space of compact operators $\mathscr{K}$.

Proof: Suppose the theorem is false. Let P be a projection of $\mathscr{B}$ onto $\mathscr{K}$. By corollary 4, there exists a projection P_0 from $\mathscr{K}$ onto a subspace $\mathscr{K}_0$, where $\mathscr{K}_0$ is isomorphic to c_0. Hence P_0P is a projection from $\mathscr{B}$ onto $\mathscr{K}_0$. This contradicts the above remark.

References

[1] Bessaga, C., and A. Pelczynski: On bases and unconditional convergence of series in Banach spaces. Studia math. 17, 150—163 (1958). — [2] Bessaga, C., and A. Pelczynski: Some remarks on conjugate spaces containing subspaces isomorphic to the space c_0. Bull. Acad. Pol. Sci. VI. 4, 249—250 (1958). — [3] Ruston, A. F.: Conjugate Banach spaces. Proc. Cambridge Phil. Soc. 5, 576—580 (1957). — [4] Schatten, R.: Completely continuous operators on a Hilbert space. Math. Ann. 134, 47—49 (1957). — [5] Schatten, R.: A theory of cross spaces. Ann. of Math., Studies No. 26, Princeton Univ. Press 1950. — [6] Zaanen, A. C.: Linear analysis. Interscience Publ. 1953.

(Received March 20, 1959)

JURCHESCU, M.
Math. Annalen, 138, 332—334 (1959)

On a theorem of Stoilow

By

MARTIN JURCHESCU in Bucarest

In connection with his definition of a covering Riemann surface STOILOW (see [4], pp. 117—119) proved the following theorem:

(A) *A (connected) 2-manifold has countable topology if it admits an interior (= open 0-dimensional) map into the sphere S^2.*

Recently GRAUERT ([2], p. 244) obtained the following statement which he used in proving that a K-complete space has countable topology.

(B) *A connected complex space has countable topology if it admits a holomorphic 0-dimensional map into a complex Euclidean space C^N.*

In this note we use the basic idea of Stoilow's proof for (A) to obtain a general theorem which connects the existence of a countable basis in a space with that of a 0-dimensional map into a space having countable topology. The theorem has purely topological character and shows in particular that the unique requirement which is decisive for the validity of (A) and (B) is that the map is 0-dimensional. Moreover we use it to obtain a simple proof of Grauert's theorem on the countability of the topology of a connected K-complete space.

Definitions. By a *space* we mean any Hausdorff space. A space is said to have *locally countable topology* if each of its points has a neighborhood having countable topology. Note that any manifold and any complex space are (locally connected and locally compact) spaces with locally countable topology.

Let X and X_0 be two spaces. A map $f: X \to X_0$ is called 0-*dimensional* if it is continuous and the fibre $f^{-1}(x_0)$ over each point $x_0 \in X_0$ has dimension $\leqq 0$.

Lemma. *Let X be a locally connected and locally compact space, X_0 any space and β_0 a basis for the topology of X_0. Then if the map $f: X \to X_0$ is 0-dimensional there is a basis in X made up of connected components of sets $f^{-1}(U_0)$ with $U_0 \in \beta_0$.*

Proof. Let x be any point in X. We have to show that for any neighborhood V of x there exists a set $U_0 \in \beta_0$ and a connected component U of $f^{-1}(U_0)$ such that $U \subset V$. In the proof of this fact we may of course assume that X is connected. Choose a neighborhood W of x in such a way that $W \subset V$, $\overline{W}$ is compact, and $B \cap f^{-1}(x_0) = 0$, where $B = \overline{W} - W$ and $x_0 = f(x)$. Such a W exists because X is locally compact and $f^{-1}(x_0)$ has dimension $\leqq 0$. Let $B_0 = f(B)$. It follows that $B_0 \cap x_0 = 0$. Moreover, since B is compact, B_0 is a closed set in X_0. Hence there exists $U_0 \in \beta_0$ such that $B_0 \cap U_0 = 0$. Let U denote the connected component containing x of $f^{-1}(U_0)$. It follows that $B \cap U = 0$, and consequently $U \subset W \cup (X - \overline{W})$. Hence, because X is connected and $W \cap U \neq 0$, we conclude that $U \subset W$, proving the lemma.

Remark. It is clear that the converse of the assertion expressed in Lemma is also true. However, we do not use it in this note.

Theorem 1. *Suppose the space X is connected, locally connected, locally compact and has locally countable topology, and the space X_0 has countable topology. Then if there exists a 0-dimensional map $f: X \to X_0$, the space X has countable topology.*

Proof. Suppose a 0-dimensional map $f: X \to X_0$ exists, and let β_0 be a countable basis for the topology of X_0. Let β be the collection of all open sets U in X for which: (1) there exists $U_0 \in \beta_0$ such that U is a connected component of $f^{-1}(U_0)$, (2) U is relatively compact, and (3) U has countable topology. We notice that (3) follows from (2) and the hypotheses about X. By Lemma, β is a basis for the topology of X. It remains to show that β is countable.

Let V be a fixed element of β, and let $\beta_{(n)}$ denote the collection of all $U \in \beta$ with the property that there is a map $i \to U_i$ from the finite set $\{1, 2, \ldots, n\}$ into β such that $U_1 = V$, $U_2 = U$ and $U_i \cap U_{i+1} \neq 0$ for $i = 1, \ldots, n-1$. Since X is connected we have

$$\beta = \bigcup_{n=1}^{\infty} \beta_{(n)} .$$

Hence it remains to show that each $\beta_{(n)}$ is countable. The assertion is manifest for $n = 1$. Now assume it is true for $n-1$, i.e. $\beta_{(n-1)}$ is countable. Let $U_0 \in \beta_0$ and $'U \in \beta_{(n-1)}$. We denote by $\alpha_{(n)}(U_0, 'U)$ the collection of all sets $U \in \beta_{(n)}$ such that U is a connected component of $f^{-1}(U_0)$ and U meets $'U$. Each set $U \in \beta_{(n)}$ meets at least one set $'U \in \beta_{(n-1)}$, so that

$$\beta_{(n)} = \bigcup_{U_0 \in \beta_0,\ 'U \in \beta_{(n-1)}} \alpha_{(n)}(U_0, 'U) .$$

Each $\alpha_{(n)}(U_0, 'U)$ is countable. For $'U$ has countable topology, and the sets $U \cap 'U$, $U \in \alpha_{(n)}(U_0, 'U)$, are not-empty, disjoint and open in $'U$. Hence $\beta_{(n)}$, as a countable union of countable collections, is countable, which completes the proof of the theorem.

As a universal space X_0 in Theorem 1 we can take the topological product $C^\omega = \prod_{i=1}^{\infty} C_{(i)}$, where each $C_{(i)}$ is the finite complex z-plane. We denote the *natural projection* $C^\omega \to C_{(i)}$ by p_i. By Theorem 1 we have

(C) *Let X be any (connected) manifold or any connected complex space. Then X has countable topology if (and only if, by Uryson's metrization theorem) it can be 0-dimensionally mapped into C^ω.*

We now use (C) to obtain a simple proof of Grauert's theorem on the countability of the topology of a connected K-complete space (see [2], pp. 243—249). We recall that a K-complete space is a complex space R having the property that for each point $x \in R$ there exists a holomorphic map of R into a complex Euclidean space C^k whose restriction to a neighborhood of x is 0-dimensional.

Theorem 2 (Grauert's theorem). *Any connected K-complete space has countable topology.*

Proof. For any complex space S we denote the set of all holomorphic functions on S by H_S and we introduce in H_S the compact (= uniform on each compact subset) convergence. We have

(1) If S has countable topology, H_S is metrizable and separable. This is an immediate consequence of Ex. 8, p. 24 in [1].

(2) If S is imbedded in a larger connected complex space R so that the set $A = R - S$ is (analytically) thin and H_S is separable metric, then H_R is separable (the statement and proof are essentially due to Grauert; see [2]).

Indeed, let us define the map $u : H_R \to H_S$ by setting $u(f) = f|S$, $f \in H_R$. Since H_S is separable metric and $u(H_R) \subset H_S$, there exists a countable dense set $\{f_i\}$ in H_R such that $\{u(f_i)\}$ is dense in $u(H_R)$. Since $A = R - S$ is thin, the convergence closure of S equals R (see [2]), proving that $\{f_i\}$ is dense in H_R.

From (1) and (2) it follows that if the connected complex space R is K-complete at a point of R, H_R is separable. For, by the definition of a K-complete space at a point and a theorem of Remmert (see [3]), there exists a holomorphic map $\tau : R \to C^k$ having the property that $\tau|S$ is 0-dimensional for an open set S in R such that $A = R - S$ is thin. By (B) (which follows from (C)), S has countable topology. From (1) it follows that H_S is separable metric, and from (2) that H_R is separable.

Now let R be any connected K-complete space, and let $\{f_i\}$ be a countable dense set in H_R. Then there exists a unique continuous map $f : R \to C^\infty$ with $f_i = p_i \circ f$. Since the property of a holomorphic map $\tau : R \to C^k$ of being not 0-dimensional in the neighborhood of a point of R is conserved by the convergence in H_R (see [2]), we see that f is 0-dimensional. Theorem 2 now follows from (C).

References

[1] Bourbaki, N.: Topologie Générale. Chap. X. Paris: Hermann et Cie 1949. — [2] Grauert, H.: Charakterisierung der holomorph vollständigen komplexen Räume. Math. Ann. **129**, 233—259 (1955). — [3] Remmert, R.: Holomorphe und meromorphe Abbildungen komplexer Räume. Math. Ann. **133**, 328—370 (1957). — [4] Stoilow, S.: Leçons sur les principes topologiques de la théorie des fonctions analytiques. Paris: Gauthier-Villars 1938 (2nd ed. 1956).

(Eingegangen am 5. März 1959)

EFFERTZ, F. H. und BREUER, K. H.
Math. Annalen 138, 335—341 (1959)

Ein Algorithmus und ein Klassifikationsprinzip für Funktionen mit nichtnegativem Realteil

Von
F. H. EFFERTZ in Krefeld und K. H. BREUER in Dortmund

Es wird ein Algorithmus für Funktionen, die in der rechten Halbebene holomorph sind und dort nichtnegativen Realteil besitzen, untersucht. Als Sonderfälle sind in ihm die Algorithmen von A. FIALKOW — I. GERST [7], R. BOTT — R. J. DUFFIN [1] und P. I. RICHARDS [10] enthalten. Durch Spezialisierung gewinnt man nach linearen Transformationen auch den bekannten Landau-Schurschen Algorithmus für beschränkte Potenzreihen [11]. Der mitgeteilte Algorithmus führt zu einer Klassifikation der Funktionen mit nichtnegativem Realteil. Die betrachteten Funktionen treten in der Theorie elektrischer Netzwerke auf und haben dann die zusätzliche Eigenschaft, daß sie auf der reellen Achse reellwertig sind [2, 5].

1. Ein Algorithmus für nichtnegative Funktionen

Es sei $F(\lambda)$ eine in der rechten Halbebene $Re\,\lambda > 0$ definierte holomorphe Funktion. Wir nennen $F(\lambda)$ eine nichtnegative Funktion, wenn sie für $Re\,\lambda > 0$ nichtnegativen Realteil besitzt. Ist $F(\lambda)$ außerdem noch auf der reellen Achse reellwertig, so sei sie nichtnegative reelle Funktion genannt. Wir beweisen nunmehr den folgenden

Satz 1: Ist $F_\nu(\lambda)$ eine nichtnegative Funktion und $\lambda_i^{(\nu)}$ $(i = 1, 2, \ldots, m_\nu)$ eine Folge von Zahlen mit $Re\,\lambda_i^{(\nu)} > 0$, für die $F_\nu(\lambda_i^{(\nu)}) = s_\nu$ $(i = 1, 2, \ldots, m_\nu)$ gilt, und ist ferner $F_\nu(\lambda)$ nicht von der Form

$$(1)\qquad F_\nu(\lambda) = \frac{s_\nu \bar{g}_\nu(\lambda) + \bar{s}_\nu \varepsilon g_\nu(-\lambda)}{\bar{g}_\nu(\lambda) - \varepsilon g_\nu(-\lambda)}$$

mit $|\varepsilon| = 1$ und dem Hurwitzpolynom

$$g_\nu(\lambda) = \prod_{i=1}^{m_\nu} (\lambda + \lambda_i^{(\nu)}),$$

dann ist auch

$$(2)\qquad F_{\nu+1}(\lambda) = \frac{K\,S_\nu(\lambda) - \bar{K}\,T_\nu(\lambda)}{S_\nu(\lambda) + T_\nu(\lambda)}$$

mit $Re\,K > 0$ und

$$(3)\qquad \begin{aligned} S_\nu(\lambda) &= \varepsilon\, g_\nu(-\lambda)\,[\bar{s}_\nu + F_\nu(\lambda)]\,,\\ T_\nu(\lambda) &= \bar{g}_\nu(\lambda)\,[s_\nu - F_\nu(\lambda)] \end{aligned}$$

eine nichtnegative Funktion.

Die auszuschließende Funktion (1) ist eine nichtnegative Funktion, deren Realteil auf der imaginären Achse außerhalb der Pole verschwindet. Setzt man $K = 1, m_\nu = 1, \lambda_1^{(\nu)} = 1$ und

$$\varepsilon = -\frac{1+s_\nu}{1+\bar{s}_\nu},$$

so folgt mit

$$\lambda = \frac{1-z}{1+z} \text{ und } F_\nu(\lambda) = \frac{1-f_\nu(z)}{1+f_\nu(z)}$$

der für die Theorie beschränkter Funktionen $f_\nu(z)$ grundlegende kettenbruchartige Reduktionsalgorithmus von J. Schur [11].

Zum Beweis von Satz 1 bilden wir mit einer unimodular beschränkten Funktion $f(z)$, für deren Nullstellen z_i $(i = 1, 2, \ldots, m)$ die Bedingung $|z_i| < 1$ gilt, die Funktion

$$\varphi(z) = f(z) \prod_{i=1}^{m} \frac{1-\bar{z}_i z}{z-z_i}.$$

Nach dem verallgemeinerten Schwarzschen Lemma gilt für $\varphi(z)$ ebenfalls $|\varphi(z)| \leq 1$ für $|z| < 1$. Durch die Transformation

$$\Phi(z) = \frac{K - \bar{K}\alpha\varphi(z)}{1+\alpha\varphi(z)}; \quad Re\, K > 0; |\alpha| = 1$$

ist eine Funktion $\Phi(z)$ mit im Einheitskreis nichtnegativem Realteil gegeben. Entsprechend wird durch

$$f(z) = \bar{\beta}\frac{s-F(\lambda)}{\bar{s}+F(\lambda)}; \quad Re\, s > 0; |\beta| = 1$$

mit

$$z = \bar{\gamma}\frac{k-\lambda}{\bar{k}+\lambda}; \quad Re\, k > 0; |\gamma| = 1 \tag{4}$$

der im Einheitskreis unimodular beschränkten Funktion $f(z)$ eine in der offenen rechten Halbebene holomorphe Funktion $F(\lambda)$ zugeordnet, die dort nichtnegativen Realteil besitzt. Ist $F(\lambda)$ eine solche nichtnegative Funktion, so ist mit $z(\lambda)$ aus (4) auch

$$F_1(\lambda) = \Phi(z(\lambda))$$

eine nichtnegative Funktion. Das Polynom

$$g(\lambda) = \prod_{i=1}^{m}(\lambda + \lambda_i)$$

ist wegen $|z_i| < 1$ und (4) ein Hurwitzpolynom, d. h. ein Polynom, dessen Nullstellen sämtlich in der offenen linken Halbebene liegen. $F_1(\lambda)$ läßt sich jetzt in der Form schreiben

$$F_1(\lambda) = \frac{K\varepsilon g(-\lambda)[\bar{s}+F(\lambda)] - \bar{K}\bar{g}(\lambda)[s-F(\lambda)]}{\varepsilon g(-\lambda)[\bar{s}+F(\lambda)] + \bar{g}(\lambda)[s-F(\lambda)]} \tag{5}$$

mit $Re\, K > 0$, $|\varepsilon| = 1$ und $s = F(\lambda_i)$ in Übereinstimmung mit (2).

Durch (5) wird jeder von (1) verschiedenen nichtnegativen Funktion wieder eine solche zugeordnet. Die einer nichtnegativen Funktion $F_0(\lambda)$ gemäß (2) zugeordneten Funktionen $F_1(\lambda), F_2(\lambda), \ldots$ heißen die zu $F_0(\lambda)$ adjungierten Funktionen und die s_ν die adjungierten Parameter.

Tritt eine Funktion von der Form (1) als adjungierte Funktion auf, so bricht der Algorithmus ab.

Satz 2: Wird Γ_ν mit $Re\,\Gamma_\nu > 0$ und ein Hurwitzpolynom $g_\nu(\lambda)$ beliebig vorgegeben und ist $F_{\nu+1}(\lambda)$ eine nichtnegative Funktion, so ist

$$F_\nu(\lambda) = \frac{\Gamma_\nu M_{\nu+1}(\lambda) - \bar{\Gamma}_\nu N_{\nu+1}(\lambda)}{M_{\nu+1}(\lambda) + N_{\nu+1}(\lambda)}$$

mit

$$M_{\nu+1}(\lambda) = \bar{g}_\nu(\lambda)\,[\bar{K} + F_{\nu+1}(\lambda)]\,, \quad Re\,K > 0$$

$$N_{\nu+1}(\lambda) = \varepsilon g_\nu(-\lambda)\,[K - F_{\nu+1}(\lambda)]\,, \quad |\varepsilon| = 1$$

eine Funktion, die für $Re\,\lambda > 0$ holomorph ist und dort positiven Realteil besitzt. Sie nimmt an den Nullstellen $\lambda_i^{(\nu)}$ von $g_\nu(-\lambda)$ den vorgegebenen Wert Γ_ν an.

Zum Beweis setzt man

$$\psi(\lambda) = \bar{\alpha}\,\frac{K - F_{\nu+1}(\lambda)}{\bar{K} + F_{\nu+1}(\lambda)}\,, \quad |\alpha| = 1, Re\,K > 0$$

und bildet

$$\Phi(\lambda) = \frac{g_\nu(-\lambda)}{\bar{g}_\nu(\lambda)}\,\psi(\lambda)\,.$$

Dann gilt $|\Phi(\lambda)| < 1$ für $Re\,\lambda > 0$. Bildet man weiter mit $\Phi(\lambda)$ die Funktion

$$F_\nu(\lambda) = \frac{\Gamma_\nu - \delta\bar{\Gamma}_\nu\Phi(\lambda)}{1 + \delta\Phi(\lambda)}\,, \quad Re\,\Gamma_\nu > 0,\ |\delta| = 1\,,$$

so ist $F_\nu(\lambda)$ für $Re\,\lambda > 0$ holomorph und hat dort positiven Realteil. Durch Einsetzen von $\Phi(\lambda)$ folgt mit $\bar{\alpha}\delta = \varepsilon$ die Behauptung.

Satz 3: Es sei $F_0(\lambda) = \dfrac{G_0(\lambda)}{H_0(\lambda)}$ mit teilerfremden Polynomen G_0 und H_0 eine rationale nichtnegative Funktion. Wird weiter der Grad von $F_0(\lambda)$ durch

$$\{F_0(\lambda)\} = \max\,[\{G_0(\lambda)\}, \{H_0(\lambda)\}]$$

bezeichnet und definiert, so gilt für den Grad der nach (2) gebildeten adjungierten Funktion $F_1(\lambda)$ nach Kürzen des gemeinsamen Faktors $g_0(-\lambda)$, wenn

$$F_0(-\bar{\lambda}_i) \neq -\overline{F_0(\lambda_i)} \qquad (i = 1, 2, \ldots, m)$$

ist:

$$\{F_1(\lambda)\} = \{F_0(\lambda)\}\,.$$

Gilt dagegen für eine Teilfolge der λ_i

$$\lambda_{i_1}, \lambda_{i_2}, \ldots, \lambda_{i_\mu}$$

die Relation

$$F_0(-\bar{\lambda}_{i_\nu}) = -\overline{F_0(\lambda_{i_\nu})} \qquad (\nu = 1, 2, \ldots, \mu)\,, \tag{6}$$

so ist

$$\{F_1(\lambda)\} = \{F_0(\lambda)\} - \mu\,.$$

Zum Beweis berücksichtige man, daß eine Gradreduktion bei einmaliger Anwendung von (2) auf $F_0(\lambda)$ dann und nur dann auftritt, wenn neben $g_0(-\lambda)$ weitere gemeinsame Faktoren vorhanden sind. (6) ist die Bedingung dafür, daß der gemeinsame Faktor

$$\prod_{\nu=1}^{\mu} (\lambda + \bar{\lambda}_{i_\nu})$$

auftritt. Eine nichtnegative Funktion hat weiter im unendlichfernen Punkt höchstens eine Null- oder Polstelle erster Ordnung, woraus dann die Behauptung folgt.

2. Ein Klassifikationsprinzip für nichtnegative Funktionen

Der Algorithmus (2) für $m_\nu = 1$ für alle ν erlaubt eine Klassifikation der nichtnegativen Funktionen in der folgenden Weise:

Klasse I: Der Algorithmus ist stets fortsetzbar, d. h. es gibt kein n, derart, daß gilt

$$S_n(\lambda) + T_n(\lambda) \equiv 0 .$$

a) $$F_n(\lambda) \not\equiv F_n(\lambda_n) , \qquad (n = 1, 2, \ldots)$$

α) $F_0(\lambda)$ rational,

β) $F_0(\lambda)$ nicht rational.

b) $$F_n(\lambda) \equiv F_n(\lambda_n).$$

Klasse II: Der Algorithmus ist nicht beliebig weit fortsetzbar, d. h. es gibt ein n, derart, daß gilt

$$S_n(\lambda) + T_n(\lambda) \equiv 0 .$$

Die für die Anwendungen in der Theorie elektrischer Netzwerke wichtigen Klassen Ib und II werden im folgenden weiter untersucht. Bei Klasse Ib lauten die weiteren adjungierten Funktionen

$$F_{n+1}(\lambda) = F_{n+2}(\lambda) = \cdots = K \quad \text{mit} \quad Re\, K > 0 .$$

Gehört die nichtnegative Funktion $F_0(\lambda)$ zur Klasse Ib oder II, so ist sie notwendig rational.

Satz 4: Die nichtkonstante rationale Funktion

$$F_0(\lambda) = \frac{G_0(\lambda)}{H_0(\lambda)}$$

mit teilerfremden Polynomen $G_0(\lambda)$ und $H_0(\lambda)$ ist dann und nur dann eine nichtnegative Funktion der Klasse Ib oder II, wenn gilt

$$\text{(7)} \quad \begin{aligned} &1)\ G_0(\lambda) + \delta H_0(\lambda) \text{ ist ein Hurwitzpolynom mit } Re\,\delta > 0 , \\ &2)\ G_0(\lambda)\, \bar{H}_0(-\lambda) + \bar{G}_0(-\lambda)\, H_0(\lambda) = C \prod_{\nu=0}^{n-1} (\lambda_\nu - \lambda)\, (\bar{\lambda}_\nu + \lambda) \end{aligned}$$

mit $Re\,\lambda_\nu > 0$ $(\nu = 0, 1, 2, \ldots, n-1)$ und $C \geqq 0$.

$C > 0$ charakterisiert die Klasse Ib und $C = 0$ die Klasse II.

Die Bedingungen (7) sind notwendig; denn ist $F_0(\lambda)$ eine rationale nichtnegative Funktion der Klasse Ib oder II und δ eine Zahl mit positivem Realteil, so ist für $Re\,\lambda > 0$

$$Re\left(\delta + \frac{G_0(\lambda)}{H_0(\lambda)}\right) = Re\,\frac{G_0(\lambda) + \delta H_0(\lambda)}{H_0(\lambda)} > 0\,,$$

also $G_0(\lambda) + \delta H_0(\lambda) \neq 0$ für $Re\,\lambda > 0$. Dies gilt auch für $Re\,\lambda = 0$, da im anderen Falle wegen der Teilerfremdheit von $G_0(\lambda)$ und $H_0(\lambda)$ und der Stetigkeit der Abbildung Punkte der rechten Halbebene auf eine Umgebung des Punktes $-\delta$ abgebildet würden.

Ferner gilt mit $F_n(\lambda) = G_n/H_n$ im Fall der Klasse Ib

$$G_n(\lambda)\,\overline{H}_n(-\lambda) + \overline{G}_n(-\lambda)\,H_n(\lambda) = 2H_n(\lambda_n)\,\overline{H_n(\lambda_n)}\,Re\,F_n(\lambda_n) = C_n > 0$$

und im Fall der Klasse II $C_n = 0$ und mit

$$F_n(\lambda) = \frac{(s_n - \varepsilon\bar{s}_n)\,\lambda + (s_n\bar{\lambda}_n + \varepsilon\bar{s}_n\lambda_n)}{(1+\varepsilon)\,\lambda + (\bar{\lambda}_n - \varepsilon\lambda_n)}$$

$$G_n(\lambda)\,\overline{H}_n(-\lambda) + \overline{G}_n(-\lambda)\,H_n(\lambda) = 0\,.$$

Bildet man zu $F_n(\lambda)$ nach Satz 2 mit $g_{n-1}(\lambda) = -(\lambda + \lambda_{n-1})$ die Funktion $F_{n-1}(\lambda)$, so erhält man

$$G_{n-1}(\lambda)\,\overline{H}_{n-1}(-\lambda) + \overline{G}_{n-1}(-\lambda)\,H_{n-1}(\lambda)$$
$$= \begin{cases} C_{n-1}(\lambda_{n-1} - \lambda)\,(\bar{\lambda}_{n-1} + \lambda) & \text{Klasse Ib} \\ 0 & \text{Klasse II}\,, \end{cases}$$

wobei $C_{n-1} = C_n(K + \overline{K})\,[\Gamma_{n-1} + \overline{\Gamma}_{n-1}]$ ist.
Durch vollständige Induktion folgt die Notwendigkeit der Bedingungen (7).

Um die Hinlänglichkeit zu beweisen, sei die Funktion

$$T(\lambda) = \frac{\bar{\delta} - F_0(\lambda)}{\delta + F_0(\lambda)} = \frac{\bar{\delta}H_0(\lambda) - G_0(\lambda)}{\delta H_0(\lambda) + G_0(\lambda)} \tag{8}$$

betrachtet. Wegen Bedingung 1) ist $T(\lambda)$ holomorph für $Re\,\lambda \geq 0$ (λ endlich). Unter Berücksichtigung der Bedingung 2) erkennt man, daß $T(\lambda)$ auch im unendlich fernen Punkt holomorph ist.

Nach Bedingung 2) ist ferner

$$|T(\lambda)| \leq 1 \quad \text{für} \quad \lambda = i\omega \quad (-\infty \leq \omega \leq \infty)\,,$$

wegen $1 - |T(i\omega)|^2 \geq 0$.

Nach dem Prinzip vom Maximum ist daher

$$|T(\lambda)| < 1 \quad \text{für} \quad Re\,\lambda > 0\,,$$

so daß aus der Abbildungseigenschaft der Transformation (8) folgt, daß $F_0(\lambda)$ eine nichtnegative Funktion ist, wenn die Bedingungen (7) gelten.

Bildet man nach Satz 1 zu $F_0(\lambda)$ die erste adjungierte Funktion $F_1(\lambda) = G_1(\lambda)/H_1(\lambda)$, so folgt aus (7) Bedingung 2) für $\lambda = -\bar{\lambda}_0$

$$G_0(-\bar{\lambda}_0)\,\overline{H_0(\lambda_0)} + \overline{G_0(\lambda_0)}\,H_0(-\bar{\lambda}_0) = 0$$

oder

$$F_0(-\bar{\lambda}_0) = -\overline{F_0(\lambda_0)}\,,$$

so daß nach (6) die adjungierte Funktion $F_1(\lambda)$ den Grad $n-1$ hat, wenn $F_0(\lambda)$ vom n-ten Grade ist. Ferner gilt nach Kürzen des gemeinsamen Faktors $(\lambda-\lambda_0)(\lambda+\bar\lambda_0)$ in Zähler und Nenner von $F_1(\lambda)$ bei Benutzung von (7):

$$G_1(\lambda)\,\bar H_1(-\lambda)+\bar G_1(-\lambda)\,H_1(\lambda)=C_1\prod_{\nu=1}^{n-1}(\lambda_\nu-\lambda)(\bar\lambda_\nu+\lambda)$$

mit

$$C_1=C(K+\bar K)(s_0+\bar s_0)\geqq 0\,.$$

Die n-malige Anwendung des Algorithmus (2) für $m_\nu=1$ auf $F_0(\lambda)$, das den Bedingungen (7) mit $C\geqq 0$ genügt, mit der durch (7) bestimmten Folge $\lambda_1^{(\nu)}=\lambda_\nu$, führt also entweder auf eine Konstante mit positivem Realteil (Klasse Ib) oder auf eine Konstante mit verschwindendem Realteil (Klasse II); oder die $(n-1)$-malige Anwendung des Algorithmus (2) führt auf ein $F_{n-1}(\lambda)$ vom ersten Grade, für das $S_{n-1}(\lambda)+T_{n-1}(\lambda)\equiv 0$ ist (Klasse II).

Beziehen wir uns im folgenden auf nichtnegative reelle Funktionen, so gestattet der hier mitgeteilte Satz 5 die ν-te adjungierte Funktion $F_\nu(\lambda)$ $(\nu=n-1,n-2,\ldots,0)$ der durch $K=1$, $\varepsilon=-1$, $\lambda_\nu=1$ $(\nu=0,1,2,\ldots,n-1)$ gegebenen Unterklasse der Klassen Ib und II ohne Bestimmung weiterer adjungierter Funktionen anzugeben. Damit ist zugleich eine explizite Darstellung der Funktionen dieser Unterklasse gewonnen.

Satz 5: Ist (A_k) $(k=n,n-1,\ldots,\nu+1,\nu)$ eine vorgegebene Folge positiver Zahlen und sind die Größen $\alpha_\nu(\lambda)$, $\beta_\nu(\lambda)$, $\gamma_\nu(\lambda)$, $\delta_\nu(\lambda)$ durch die Matrizengleichung

$$(9)\qquad \begin{Vmatrix}\alpha_\nu & \beta_\nu\\ \gamma_\nu & \delta_\nu\end{Vmatrix}=\begin{Vmatrix}A_\nu & A_\nu\lambda\\ \lambda & 1\end{Vmatrix}\cdot\begin{Vmatrix}A_{\nu+1} & A_{\nu+1}\lambda\\ \lambda & 1\end{Vmatrix}\cdots\begin{Vmatrix}A_{n-1} & A_{n-1}\lambda\\ \lambda & 1\end{Vmatrix}\begin{Vmatrix}A_n & 0\\ 0 & 1\end{Vmatrix}$$

gegeben, so gewinnt man die adjungierten Funktionen der genannten Unterklasse der Klasse Ib aus

$$(10\text{a})\qquad F_\nu(\lambda)=\frac{\alpha_\nu+\beta_\nu}{\gamma_\nu+\delta_\nu}\qquad(\nu=n-1,n-2,\ldots,0)\,,$$

$$(10\text{b})\qquad F_n(\lambda)=A_n\,.$$

Die entsprechenden adjungierten Funktionen der Klasse II erhält man erstens aus

$$(11)\qquad \begin{Vmatrix}\alpha_\nu & \beta_\nu\\ \gamma_\nu & \delta_\nu\end{Vmatrix}=\begin{Vmatrix}A_\nu & A_\nu\lambda\\ \lambda & 1\end{Vmatrix}\cdot\begin{Vmatrix}A_{\nu+1} & A_{\nu+1}\lambda\\ \lambda & 1\end{Vmatrix}\cdots\begin{Vmatrix}A_{n-1} & A_{n-1}\lambda\\ \lambda & 1\end{Vmatrix}\cdot\begin{Vmatrix}A_n & 0\\ 0 & \lambda\end{Vmatrix}$$

mit (10a), statt (10b) gilt jetzt $F_n(\lambda)=\frac{A_n}{\lambda}$. Die restlichen Funktionen der Klasse II erhält man mit (9) aus (10a) und (10b) für $A_s>0$ $(s=n-1,n-2,\ldots,\nu+1,\nu)$, $A_n=0$. Der Beweis folgt leicht durch vollständige Induktion. Die angegebenen Matrizendarstellungen charakterisieren explizit Klassen von Frequenzcharakteristiken elektrischer Netzwerke. Abschließend sei noch bemerkt, daß die nichtnegativen Funktionen der Form

$$(12)\qquad F(\lambda)=\frac{\bar\delta\,h(\lambda)-\delta\,\varepsilon\,\bar h(-\lambda)}{h(\lambda)+\varepsilon\,\bar h(-\lambda)}\,,$$

wobei $h(\lambda)$ ein Hurwitzpolynom ist und $Re\,\delta>0$, $|\varepsilon|=1$ gilt, sämtlich zur Klasse II gehören [6]. Für den Sonderfall $\varepsilon=1$ und $h(\lambda)=\bar h(\lambda)$ erhält man

einen Satz von W. Cauer [3], der einen Beweis einer charakteristischen Parameterdarstellung der Koeffizienten reeller Hurwitzpolynome ermöglicht [4].

Aus der Darstellung (12) folgt, daß die Wurzeln der Polynome

$$f_1(\lambda) = h(\lambda) - \varepsilon \bar{h}(-\lambda) \text{ und}$$
$$f_2(\lambda) = h(\lambda) + \varepsilon \bar{h}(-\lambda)$$

auf der imaginären Achse liegen, sämtlich voneinander verschieden sind und sich trennen. Entsprechend gilt die folgende Bemerkung, die einen Satz von J. Schur ([11], Satz XII) verallgemeinert:

Ist $P(\lambda)$ ein Polynom, dessen sämtliche Nullstellen außerhalb des Einheitskreises liegen, so liegen die Wurzeln der Polynome

$$P_1(\lambda) = P(\lambda) + \varepsilon \lambda^n P(\lambda^{-1}) \text{ und}$$
$$P_2(\lambda) = P(\lambda) - \varepsilon \lambda^n P(\lambda^{-1})$$

auf dem Einheitskreis, sind sämtlich voneinander verschieden und trennen sich.

Aus dieser Bemerkung folgt mit Benutzung eines Theorems von C. F. Gauss [8] der folgende Satz[1]):

Sind die Wurzeln von $F(\lambda) = \sum_{\nu=0}^{n} c_\nu \lambda^\nu$ voneinander verschieden und auf dem Einheitskreis gelegen, so trennen sich die sämtlich auf dem Einheitskreis gelegenen und sämtlich einfachen Wurzeln von

$$\hat{P}_1(\lambda) = (1-\varepsilon)\,\lambda F'(\lambda) - n\,F(\lambda) \text{ und}$$
$$\hat{P}_2(\lambda) = (1+\varepsilon)\,\lambda F'(\lambda) - n\,F(\lambda)\,.$$

Aus einem Satz von Laguerre [9] folgt der folgende Sonderfall des letzten Satzes:

Liegen die Wurzeln von $F(\lambda)$ sämtlich auf dem Einheitskreis, so gilt dies auch für die Wurzeln von

$$\hat{P}_2(\lambda) = 2\lambda F'(\lambda) - n\,F(\lambda)\,.$$

Literatur

[1] Bott, R., and R. J. Duffin: Impedance Synthesis Without Use of Transformers. J. appl. Physics 20, 816 (1949). — [2] Cauer, W.: Über Funktionen mit positivem Realteil. Math. Ann. 106, 369 (1932). — [3] Cauer, W.: Frequenzweichen konstanten Betriebswiderstandes. Elektr. Nachr. Techn. 16, 96 (1939), Anhang I. — [4] Cremer, H., u. F. H. Effertz: Über die algebraischen Kriterien für die Stabilität von Regelungssystemen. Math. Ann. 137, 328 (1959). — [5] Effertz, F. H.: On the Synthesis of Networks Containing Two Kinds of Elements. Proc. Symp. Mod. Netw. Synthesis 5, 145 (1955). — [6] Effertz, F. H.: Beschränkte Funktionen, Frequenzcharakteristiken elektrischer Netzwerke und algebraische Stabilitätskriterien. Z. angew. Math. Mech. 33, 281 (1953). — [7] Fialkow, A., and I. Gerst: Impedance Synthesis Without Minimization. J. Math. Phys. 34, 160 (1955). — [8] Gauss, C. F.: Werke III, 112. — [9] Polya-Szegö, G.: Aufgaben und Lehrsätze aus der Analysis. Bd. II, S. 58. Berlin: Springer 1954. — [10] Richards, P. I.: A Special Class of Functions with Positive Real Parts in a Half-Plane. Duke Math. J. 14, 777 (1947). — [11] Schur, J.: Über Potenzreihen, die im Innern des Einheitskreises beschränkt sind. J. f. Math. 147, 205 (1917); 148, 122 (1918).

(Eingegangen am 10. März 1959)

[1]) Herrn Dozent Dr. Prachar in Wien verdanken wir einen anderen Beweis dieses Satzes.

RAMSPOTT, K.-J.
Math. Annalen 138, 342—355 (1959)

Existenz von Holomorphiegebieten zu vorgegebener erster Bettischer Gruppe

Von
KARL-JOSEF RAMSPOTT in München

Einleitung

Im Jahre 1953 hat J. P. SERRE ([11], Corollaire 2) gezeigt, daß für jede holomorph vollständige Mannigfaltigkeit M^n der komplexen Dimension n die ganzzahligen Bettischen Gruppen $B_\nu(M^n)$ stets nur das Nullelement enthalten, falls $\nu > n$ ist[1]). Über die Struktur der Bettischen Gruppen $B_\nu(M^n)$ mit $\nu \leqq n$ weiß man jedoch fast nichts. Es liegt nahe, sich die Frage vorzulegen, ob man zu vorgegebenen abzählbaren, torsionsfreien abelschen Gruppen $B_1, \ldots, B_n$ stets ein Holomorphiegebiet G im C^n $(n > 1)$ finden kann, dessen ν-te Bettische Gruppe $B_\nu(G)$ zur Gruppe B_ν isomorph ist $(\nu = 1, \ldots, n)$.

In dieser Arbeit wird ein spezieller Beitrag zu diesem Problem gegeben. Es wird bewiesen:

Ist B eine beliebige abzählbare, torsionsfreie abelsche Gruppe, so gibt es stets ein Holomorphiegebiet G im C^n $(n > 1)$, dessen erste Bettische Gruppe $B_1(G)$ zu B isomorph ist[2]).

Zur Vorbereitung wird im ersten Paragraphen neben einer komplexen Mannigfaltigkeit M^n noch die komplexe Mannigfaltigkeit $M^n - A^p$ betrachtet, dabei ist A^p eine rein p-dimensionale analytische Menge in M^n. Es werden einige einfache Beziehungen zwischen den ganzzahligen Homologiegruppen beider Mannigfaltigkeiten angegeben. Ferner wird ein für die Konstruktion von G benötigter gruppentheoretischer Hilfssatz (vgl. § 1.3) bewiesen.

Im zweiten Paragraphen wird die Existenz eines gesuchten Holomorphiegebietes nachgewiesen. Im zweiten Abschnitt wird die Idee erläutert, die dem Beweis zugrunde liegt. Im Prinzip handelt es sich dabei um eine Methode, die schon L. PONTRJAGIN in ([7], Anhang III) benutzt hat, um eine Kurve im R^3 anzugeben, deren Komplementärraum eine beliebige abzählbare, torsionsfreie abelsche Gruppe als erste Bettische Gruppe hat. K. STEIN hat in [12] dieses Verfahren in entsprechend modifizierter Form benutzt, um im C^2 ein Holomorphiegebiet zu bestimmen, dessen erste Bettische Gruppe zur additiven Gruppe der rationalen Zahlen von der Gestalt $\frac{q}{2^\mu}$ (q, μ ganz) isomorph ist.

Es sei mir gestattet, an dieser Stelle Herrn Professor Dr. K. STEIN für wertvolle Anregungen meinen herzlichen Dank auszusprechen.

[1]) Unter der q-ten Homologiegruppe $H_q(M^n)$ wird wie üblich die Gruppe der Homologieklassen der in M^n gelegenen singulären, ganzzahligen q-dimensionalen Zyklen verstanden. Die Faktorgruppe nach dem Torsionsanteil wird als q-te Bettische Gruppe $B_q(M^n)$ bezeichnet (vgl. etwa [10]).

[2]) Die erste Bettische Gruppe eines Gebietes der komplexen Ebene ist stets frei.

§ 1. Vorbereitende Aussagen

1. M^n sei eine zusammenhängende komplex-analytische Mannigfaltigkeit der komplexen Dimension n, A^p, $0 \leq p < n$, sei eine in M^n gelegene rein p-dimensionale analytische Menge[3]). Dann ist der natürliche Homomorphismus der q-ten Homologiegruppe $H_q(M^n - A^p)$ in die q-te Homologiegruppe $H_q(M^n)$ für $q \leq 2n - 2p - 2$ ein Isomorphismus:

$$(*) \qquad H_q(M^n - A^p) \cong H_q(M^n)\,, \quad q \leq 2n - 2p - 2.$$

Das folgt unmittelbar aus der exakten Homologiesequenz

$$(**) \qquad \to H_{q+1}(M^n) \to H_{q+1}(M^n \bmod M^n - A^p) \to H_q(M^n - A^p) \to H_q(M^n) \to,$$

wenn man die Alexander-Pontrjaginsche Dualität[4]) ausnutzt. Danach ist

$$H_q(M^n \bmod M^n - A^p) \cong H_*^{2n-q}(A^p)\,,$$

wenn $H_*^{2n-q}(A^p)$ die $(2n-q)$-te Cohomologiegruppe von A^p mit kompaktem Träger und dem Ring der ganzen Zahlen als Koeffizientenbereich bezeichnet. Da $H_*^{2n-q-1}(A^p)$ verschwindet für $2n - q - 1 > 2p$, ist $(*)$ bewiesen.

2. Im folgenden sei die analytische Menge A stets rein $(n-1)$-dimensional. Außerdem sei sie zunächst als irreduzibel vorausgesetzt. Dann folgt aus $(**)$ nur, daß der natürliche Homomorphismus $H_1(M^n - A^{n-1}) \to H_1(M^n)$ ein Epimorphismus ist. Um den Kern dieses Homomorphismus zu bestimmen, entferne man aus M^n die Menge S der singulären Punkte von A^{n-1}. Es sei $'M^n := M^n - S$ und $'A^{n-1} := A^{n-1} - S$. $'A^{n-1}$ ist zusammenhängend und als Mannigfaltigkeit in $'M^n$ eingebettet. Dann folgt:

$$\to H_2('M^n) \to H_2('M^n \bmod 'M^n - 'A^{n-1}) \to H_1('M^n - 'A^{n-1}) \to H_1('M^n) \to$$

ist exakt. Da die Dimension von S höchstens $2n-4$ ist, gilt wegen $(*)$ $H_\nu('M^n) = H_\nu(M^n)$, $\nu = 1, 2$. Da ferner $H_*^{2n-1}('A^{n-1})$ verschwindet, bekommt man die exakte Sequenz

$$\to H_2(M^n) \to H_*^{2n-2}('A^{n-1}) \to H_1(M^n - A^{n-1}) \to H_1(M^n) \to 0\,.$$

$'A^{n-1}$ ist Mannigfaltigkeit. Daher hat man für komplementäre Dimensionen die Poincarésche Dualität[5]) zwischen der Cohomologie mit kompaktem Träger und der Homologie der endlichen singulären Ketten. Da ferner $'A^{n-1}$ zusammenhängend ist, bekommt man die Isomorphien $H_*^{2n-2}('A^{n-1}) \cong H_0('A^{n-1}) \cong Z$, dabei ist Z die additive Gruppe der ganzen Zahlen. Damit ergibt sich die exakte Sequenz

$$\to H_2(M^n) \xrightarrow{j} Z \to H_1(M^n - A^{n-1}) \to H_1(M^n) \to 0\,.$$

Die Quotientengruppe $Z_m := Z/j(H_2(M^n))$ ist zyklisch; ihre Ordnung sei m. Dann gilt:

$$0 \to Z_m \to H_1(M^n - A^{n-1}) \to H_1(M^n) \to 0$$

ist exakt.

---

[3]) Zur Theorie der analytischen Mengen vgl. z. B. [8], [9].

[4]) Vgl. [4], Exp. XVII, ferner [5], p. 113.

[5]) Vgl. [4], Exp. XVII.

Die nicht negative ganze Zahl m hat eine einfache topologische Bedeutung. Es sei zunächst $m \neq 0$. Dann werden diejenigen 1-dimensionalen Zyklen in $M^n - A^{n-1}$, die in M^n nullhomolog sind, erzeugt durch einen Zyklus α^1, dessen m-faches bereits in $M^n - A^{n-1}$ nullhomolog ist: $m\alpha^1 \sim 0$ in $M^n - A^{n-1}$. m ist die kleinste natürliche Zahl mit dieser Eigenschaft. Ist $z^{(0)}$ ein gewöhnlicher Punkt von A^{n-1}, so gibt es eine Umgebung $U(z^{(0)})$, in der lokale Koordinaten $z_1, \ldots, z_n$ so eingeführt werden können, daß $U(z^{(0)}) \cap A^{n-1}$ durch $z_n = 0$ gegeben wird. Dann darf angenommen werden, daß der Zyklus α^1 in $U(z^{(0)})$ liegt und auf der Ebene $E^1: \{z_1 = 0, \ldots, z_{n-1} = 0\}$ durch $z_n = \varrho e^{i\varphi}$, $0 \leq \varphi \leq 2\pi$, bestimmt ist. Es sei $\varkappa^2$ die von α^1 auf E^1 berandete, in natürlicher Weise orientierte Kreisscheibe. Dann ist $m\alpha^1 = \delta m\varkappa^2$ in M^n. Weiter gibt es in $M^n - A^{n-1}$ eine 2-dimensionale Kette β^2, so daß $-m\alpha^1 = \delta\beta^2$ in $M^n - A^{n-1}$ gilt. Folglich ist $\sigma^2 = \beta^2 + m\varkappa^2$ ein 2-dimensionaler Zyklus in M^n, der mit A^{n-1} genau die Schnittzahl[6]) m hat. Verschwinden aber sämtliche Schnittzahlen von A^{n-1} mit den Elementen von $H_2(M^n)$, so muß der Kern des Homomorphismus $H_1(M^n - A^{n-1}) \to H_1(M^n)$ notwendig durch einen Zyklus α^1 erzeugt werden, für den gilt: $\nu\alpha^1 \nsim 0$ in $M^n - A^{n-1}$ für alle natürlichen Zahlen ν.

Damit ist offenbar die folgende Aussage bewiesen:

Satz 1: *M^n sei eine zusammenhängende komplexe Mannigfaltigkeit, A^{n-1} eine irreduzible, $(n-1)$-dimensionale analytische Menge in M^n. Dann gilt für die ersten Homologiegruppen $H_1(M^n)$ und $H_1(M^n - A^{n-1})$ die Isomorphie*

$$H_1(M^n) \cong H_1(M^n - A^{n-1})/Z_m\,.^{7)}$$

Verschwinden sämtliche Schnittzahlen der Menge A^{n-1} mit den Elementen von $H_2(M^n)$, so ist Z_m durch Z zu ersetzen, andernfalls ist m die kleinste von Null verschiedene positive Schnittzahl von A^{n-1} mit den Elementen von $H_2(M^n)$.

Als unmittelbare Folgerung hat man

Satz 2: *M^n sei eine zusammenhängende komplexe Mannigfaltigkeit, $H_1(M^n)$ sei frei. Ferner seien A^{n-1}_ϱ, $\varrho = 1, \ldots, r$, irreduzible, $(n-1)$-dimensionale analytische Mengen in M^n, deren Schnittzahlen mit den Elementen von $H_2(M^n)$ sämtlich verschwinden. Dann lassen sich $H_1(M^n - A^{n-1})$ und $B_1(M^n - A^{n-1})$ als direkte Summen darstellen:*

$$H_1(M^n - A^{n-1}) \cong H_1(M^n) \oplus J$$
$$B_1(M^n - A^{n-1}) \cong B_1(M^n) \oplus J\,.$$

Dabei ist $A^{n-1} := \bigcup_{\varrho=1}^{r} A^{n-1}_\varrho$ und J ein r-dimensionales Gitter[8]).

---

[6]) Zum Begriff der Schnittzahl vgl. etwa [1], [10]. In einer orientierten Mannigfaltigkeit ist die Schnittzahl $s(F_1, F_2)$ zweier abgeschlossener orientierter Flächen F_1 und F_2 von komplementärer Dimension genau dann definiert, wenn gilt: $\delta F_1 \cap F_2 = F_1 \cap \delta F_2 =$ leer (Schnittzahlbedingung). Vgl. [6], Fußnote 13.

[7]) Mit Z werde die additive Gruppe der ganzen Zahlen, mit Z_m die additive Gruppe der ganzen Zahlen modulo m bezeichnet. Speziell ist $Z_0 = Z$ und Z_1 die Nullgruppe.

[8]) Eine abelsche Gruppe mit einer endlichen Basis aus r Elementen (jede andere Basis hat dann dieselbe Länge) heißt auch ein r-dimensionales Gitter.

Beim Beweis hat man nur zu berücksichtigen, daß $A^{n-1}_{\varrho_0}$ auch mit den 2-dimensionalen Zyklen von $M^n_{\varrho_0} := M^n - (A^{n-1}_1 \cup \cdots \cup A^{n-1}_{\varrho_0-1})$, $1 < \varrho_0 \leq r$, die Schnittzahl Null hat, da $A^{n-1}_{\varrho_0}$ in ganz M^n analytisch ist.

Die Homologiegruppen stimmen mit den Bettischen Gruppen überein, da $H_1(M^n)$ nach Voraussetzung keine Elemente endlicher Ordnung besitzt.

3. Für die Konstruktion des gewünschten Holomorphiegebietes benötigen wir noch einen gruppentheoretischen Hilfssatz. Bekanntlich gilt für jede endlich erzeugte abelsche Gruppe die Aussage:

Jede abelsche Gruppe mit endlich vielen Erzeugenden ist eine direkte Summe $B = J \oplus T$, wobei T die Gruppe aller Elemente endlicher Ordnung von B und J ein r-dimensionales Gitter mit $r = \text{Rang}\,(B)$ ist[9]).

Als triviale Folgerung hat man:

Eine abelsche Gruppe mit endlich vielen Erzeugenden, die kein von Null verschiedenes Element endlicher Ordnung enthält, ist ein Gitter.

Im folgenden werden im wesentlichen nur abelsche Gruppen mit unendlich vielen Erzeugenden betrachtet. B wird stets als abzählbar vorausgesetzt. Hauptaufgabe wird sein, torsionsfreie, abzählbare abelsche Gruppen, die nicht frei sind, in ganz bestimmter Weise durch freie abelsche Gruppen auszuschöpfen.

Definition: *Eine Folge von Untergruppen B_ν $(\nu = 1, 2, \ldots)$ einer abzählbaren abelschen Gruppe B heißt eine Ausschöpfung von B, wenn*

(1) *für jedes ν die Gruppe B_ν in $B_{\nu+1}$ enthalten ist,*

(2) *B gleich der Vereinigung der B_ν ist.*

Setzen wir voraus, daß alle Untergruppen B_ν einer Ausschöpfung von B freie abelsche Gruppen vom Rang r sind, so läßt sich die Einbettung von B_ν in $B_{\nu+1}$ durch eine quadratische Matrix beschreiben. Ist $\mathfrak{t}_\nu = \begin{pmatrix} t_{\nu 1} \\ \vdots \\ t_{\nu r} \end{pmatrix}$ eine Basis von B_ν, dann ist $\mathfrak{t}_\nu = \mathfrak{A}_{\nu+1}\mathfrak{t}_{\nu+1}$; dabei ist $\mathfrak{A}_{\nu+1} = (\lambda^{(\nu+1)}_{\sigma\varrho})$ eine r-reihige, nicht singuläre, ganzzahlige quadratische Matrix. Mit dieser Bezeichnung gilt

Hilfssatz 1: *Es sei B eine torsionsfreie, abzählbare abelsche Gruppe vom Rang r. Dann existiert stets eine Ausschöpfung von B durch r-dimensionale Gitter B_ν derart, daß für die Elemente $\lambda^{(\nu+1)}_{\sigma\varrho}$ der die Einbettung von B_ν in $B_{\nu+1}$ vermittelnden ganzzahligen Matrix $\mathfrak{A}_{\nu+1}$ gilt:*

$$\lambda^{(\nu+1)}_{\sigma\varrho} \begin{cases} = 0 \text{ für } \sigma < \varrho \\ > 0 \text{ für } \sigma \geq \varrho \end{cases}, \quad \begin{matrix} \sigma = 1, \ldots, r \\ \varrho = 1, \ldots, r. \end{matrix}$$

Beweis: Wir können annehmen, daß B nicht frei ist, sonst ist nichts zu beweisen. Seien $t_1, \ldots, t_r$ r linear unabhängige Elemente in B und B_1 die von ihnen erzeugte Untergruppe. B_1 ist ein r-dimensionales Gitter. Sodann gebe man sich eine Abzählung I sämtlicher Elemente von B vor, beginnend mit den

[9]) Lineare Abhängigkeit in einer abelschen Gruppe B ist wie üblich definiert. Wenn es in B ein System von r linear unabhängigen (d. h. nicht linear abhängigen) Elementen, aber kein System von mehr als r linear unabhängigen Elementen gibt, so heißt die Zahl r der Rang von B, $r = \text{Rang}\,(B)$. Wenn es keine Zahl r mit dieser Maximaleigenschaft gibt, so sagt man, B habe den Rang unendlich. Vgl. [1], Anhang I.

Basiselementen $t_1, \ldots, t_r$ der Untergruppe B_1. Die r-dimensionalen Gitter B_μ seien für $\mu \leqq \nu$ bereits konstruiert, und es gelte $B_{\mu-1} \subset B_\mu$. Dann streiche man in dem System I sämtliche Elemente der Untergruppe B_ν. t sei das erste nicht gestrichene Element. Da B nicht frei ist, gibt es ein solches Element. Die Basiselemente von B_ν und das Element t erzeugen eine B_ν umfassende freie abelsche Gruppe $B_{\nu+1}$, die ebenfalls den Rang r hat. Damit ist eine nicht abbrechende Folge von freien abelschen Untergruppen B_ν, die sämtlich den Rang r haben, bestimmt. Ist $t \in B$ ein beliebiges Element, so ist es in der Abzählung I nach endlich vielen Schritten gestrichen, d. h. es gibt eine natürliche Zahl μ, so daß $t \in B_\mu$. Die Folge $\{B_\nu\}$ bildet also eine Ausschöpfung von B. Durch Anwendung unimodularer Transformationen werden sodann sukzessive in den B_ν, $\nu = 2, 3, \ldots$, neue Basiselemente eingeführt, so daß bezüglich dieser neuen Basen die Elemente der Übergangsmatrizen $\mathfrak{A}_{\nu+1}$ nicht negativ sind. Dazu hat man nötigenfalls die vermöge der Ausschöpfung gegebenen Übergangsmatrizen einer endlichen Anzahl von elementaren Spaltenumformungen zu unterwerfen. Man geht zweckmäßigerweise so vor: Zunächst werden die $\mathfrak{A}_{\nu+1}$ auf Dreiecksform gebracht (oberhalb der Hauptdiagonalen stehen nur Nullen), sodann werden durch Vorzeichenwechsel sämtliche Elemente der Hauptdiagonalen positiv gemacht. Die Elemente der ersten σ Zeilen in und links der Hauptdiagonalen seien bereits positiv (die Annahme ist richtig für $\sigma = 1$). Durch Addition geeigneter Vielfachen der $(\sigma + 1)$-ten Spalte zu den vorangehenden Spalten läßt sich dann jedes Element der $(\sigma + 1)$-ten Zeile positiv machen. Dabei bleiben die σ ersten Zeilen ungeändert, da die $(\sigma + 1)$-te Spalte oberhalb der Hauptdiagonalen nur aus Nullen besteht. Damit ist Hilfssatz 1 bewiesen.

§ 2. Nachweis der Existenz

1. Wir sind jetzt in der Lage, den Hauptsatz dieser Arbeit zu beweisen.

Satz 3: *Zu jeder beliebig vorgegebenen abzählbaren, torsionsfreien abelschen Gruppe B existiert stets ein Holomorphiegebiet G im C^n $(n > 1)$, dessen erste Bettische Gruppe $B_1(G)$ zur vorgegebenen Gruppe B isomorph ist.*

Ist B eine freie abelsche Gruppe vom Rang r, so ist die Konstruktion von G sehr einfach: Man entferne aus dem Polyzylinder $P: \{|z_1| < 1, \ldots, |z_n| < 1\}$ r $(n-1)$-dimensionale analytische Ebenen $E_\varrho: \{z_1 = z_1^{(\varrho)}, |z_\nu| < 1\}$, $\nu = 1, \ldots, n$; $\varrho = 1, \ldots, r$. $G := P - \bigcup\limits_{\varrho} E_\varrho$ ist Holomorphiegebiet und hat nach Satz 2 die gewünschte topologische Eigenschaft. Man kann für $B_1(G)$ auch direkt eine Basis aus r Elementen angeben. Ist B frei, aber von unendlichem Rang, so entferne man aus P eine Folge von analytischen Ebenen $E_\mu: \{z_1 = z_1^{(\mu)}, |z_\nu| < 1\}$, $\mu = 1, 2, \ldots$; dabei möge die Folge $\{z_1^{(\mu)}\}$ in der z_1-Ebene gegen einen Randpunkt des Kreises $|z_1| < 1$ konvergieren. $G := P - \bigcup\limits_{\mu} E$ ist Holomorphiegebiet. Für seine erste Bettische Gruppe läßt sich sofort eine abzählbar unendliche Basis angeben.

2. Wir werden für den Rest des Beweises von Satz 3 die abelsche Gruppe B stets als nicht frei voraussetzen. Die Idee, auf die sich die Konstruktion von G

stützt, läßt sich wie folgt beschreiben. Man schöpft die vorgegebene abelsche Gruppe B gemäß Hilfssatz 1 durch eine aufsteigende Folge von freien abelschen Untergruppen B_ν aus. Zu dieser Folge von freien abelschen Gruppen B_ν wird sodann eine Folge von Holomorphiegebieten G_ν konstruiert, derart, daß $B_1(G_\nu)$ zu B_ν isomorph ist. Die Konstruktion der Folge $\{G_\nu\}$ geschieht so, daß die G_ν gegen ein Grenzgebiet G konvergieren, dessen 1. Bettische Gruppe $B_1(G)$ zur vorgegebenen Gruppe B isomorph ist. Die Holomorphiegebiete G_ν bekommt man, indem man etwa aus dem Dizylinder[10]) $D: \{|z_1| < 1, |z_2| < 1\}$ eine rein 1-dimensionale analytische Menge F_ν herausnimmt. Die Zahl der irreduziblen Komponenten von F_ν ist dabei gleich dem Rang von B_ν. Im nächsten Schritt ersetzt man F_ν durch die 1-dimensionale analytische Menge $F_{\nu+1}$. Es ist $G_{\nu+1} := D - F_{\nu+1}$. Dabei ist entscheidend, daß die irreduziblen Komponenten von $F_{\nu+1}$ die irreduziblen Komponenten von F_ν mit wohlbestimmter Vielfachheit approximieren.

Es ist zweckmäßig, der Konstruktion einen einfachen Hilfssatz vorauszuschicken.

Hilfssatz 2: *Es seien $\psi(z_1, z_2) = z_2^n + a_{n-1}(z_1)z_2^{n-1} + \cdots + a_0(z_1)$ ein Polynom in z_1 und z_2 ohne mehrfache Faktoren, M das Nullstellengebilde von $\psi(z_1, z_2)$ und D der Dizylinder $\{|z_1| < 1, |z_2| < 1\}$. Der Durchschnitt $A := D \cap M$ sei nicht leer, der Durchschnitt von M mit der Menge $S: \{|z_1| \leq 1, |z_2| = 1\}$ sei leer. Dann gibt es zu jeder ε^*-Umgebung*

$$U_{\varepsilon^*}(A) := \{(z_1, z_2) \in D : |\psi(z_1, z_2)| < \varepsilon^*\}$$

eine in $U_{\varepsilon^*}(A)$ enthaltene Umgebung $U(A)$ mit folgenden Eigenschaften:

(1) *Es gibt ein positives $\varepsilon < \varepsilon^*$, so daß $U_\varepsilon(A)$ in $U(A)$ enthalten ist.*

(2) *Der Durchschnitt des Ebenenstücks $E(z_1^{(0)}): \{z_1 = z_1^{(0)}, |z_2| < 1\}$, $|z_1^{(0)}| < 1$, mit $U(A)$ besteht aus höchstens n punktfremden Kreisscheiben.*

(3) *Der natürliche Homomorphismus $H_1(D - U(A)) \to H_1(D - A)$ ist ein Isomorphismus.*

Beweis: Es gibt im abgeschlossenen Dizylinder $\bar{D}$ höchstens endlich viele Punkte, die gleichzeitig Nullstellen von $\psi(z_1, z_2)$ und $\frac{\partial \psi(z_1, z_2)}{\partial z_2}$ sind. Ihre z_1-Koordinaten seien $z_1^{(1)}, \ldots, z_1^{(l)}$. Um die Schnittpunkte $(z_1^{(\lambda)}, z_{\lambda 2}^{(1)}), \ldots, (z_1^{(\lambda)}, z_{\lambda 2}^{(k_\lambda)})$ des Ebenenstücks $\{z_1 = z_1^{(\lambda)}, |z_2| < 1\}$, $\lambda = 1, \ldots, l$, mit M legen wir kleine abgeschlossene Dizylinder $K_{\lambda\varkappa}: \{|z_2 - z_{\lambda 2}^{(\varkappa)}| \leq \varepsilon_\lambda, |z_1 - z_1^{(\lambda)}| \leq \delta(\varepsilon_\lambda)\}$. Dabei seien die $\varepsilon_\lambda > 0$ so gewählt, daß die Kreise $|z_2 - z_{\lambda 2}^{(\varkappa)}| \leq \varepsilon_\lambda$, $\varkappa = 1, \ldots, k_\lambda$, in der z_2-Ebene punktfremd sind. Die $\delta(\varepsilon_\lambda) > 0$ seien so bestimmt, daß erstens die Kreise $|z_1 - z_1^{(\lambda)}| \leq \delta(\varepsilon_\lambda)$ in der z_1-Ebene punktfremd sind und zweitens die Wurzeln von $\psi = 0$ für z_1 aus $|z_1 - z_1^{(\lambda)}| \leq \delta(\varepsilon_\lambda)$ in die Kreise $|z_2 - z_{\lambda 2}^{(\varkappa)}| \leq \frac{1}{2}\varepsilon_\lambda$ fallen. Es sei $K := \bigcup (D \cap K_{\lambda\varkappa})$. Dann ist $D^* := D - K$

[10]) Es genügt, das Gebiet G im C^2 zu konstruieren. Das Holomorphiegebiet $G \times H$ mit $H: \{|z_3| < 1, \ldots, |z_n| < 1\}$ hat nämlich die gewünschte Eigenschaft im C^n ($n > 2$): $B_1(G \times H) \cong B_1(G) \cong B$.

ein Gebiet mit trivialer 1. Homologiegruppe. Es sei $A^*: = D^* \cap A$. Die analytischen Mengen A bzw. A^* zerfallen in D bzw. D^* in die gleiche Anzahl von irreduziblen Komponenten. Nach Satz 2 ist $H_1(D - A) \cong H_1(D^* - A^*)$. T sei das Komplement der Kreise $|z_1 - z_1^{(\lambda)}| < \delta(\varepsilon_\lambda)$ bezüglich $|z_1| \leq 1$. Die Schnittpunkte von A mit dem Ebenenstück $\{z_1 = z_1^{(0)}, |z_2| < 1\}$, $z_1^{(0)} \in T$, sind sämtlich einfach. $d(z_1^{(0)})$ sei die Minimaldistanz zweier solcher Punkte und $r(z_1^{(0)})$ die Minimaldistanz zur Menge S. Dann sind $d_0: = \min\limits_{z_1^{(0)} \in T} d(z_1^{(0)})$ und $r_0: = \min\limits_{z_1^{(0)} \in T} r(z_1^{(0)})$ beide positiv. Es sei $\varrho_0 > 0$ und $\varrho_0 < \varrho'_0: = \min\limits_{\lambda = 1, \ldots, l} \left(r_0, \frac{d_0}{2}, \frac{\varepsilon_\lambda}{2}\right)$. Wir legen um jeden Punkt $z^{(0)} \in A^*$ auf der Ebene $z_1 = z_1^{(0)}$ die Kreisscheibe $V(z^{(0)}): |z_2 - z_2^{(0)}| < \varrho_0$. Es sei $V: = \bigcup\limits_{z^{(0)} \in A^*} V(z^{(0)})$. Dann läßt sich die identische Abbildung i_0 von $D^* - A^*$ homotop in eine stetige surjektive Abbildung $i_1: D^* - A^* \to D^* - V$ deformieren. Es sei $U(A): = V \cup K$, dann ist $D^* - V = D - U(A)$, also $H_1(D - A) \cong H_1(D^* - A^*) \cong H_1(D^* - V) = H_1(D - U(A))$. Es ist klar, daß $U(A)$ ein $U_\varepsilon(A): |\psi(z_1, z_2)| < \varepsilon$, $\varepsilon > 0$, enthält. Umgekehrt kann man bei Vorgabe von $U_{\varepsilon^*}(A)$ stets ein $U(A)$ mit den gewünschten Eigenschaften finden, so daß $U(A) \subset U_{\varepsilon^*}(A)$. Man hat nur die $K_{1\varkappa}$ und ϱ_0 hinreichend klein zu wählen.

3. Zur Konstruktion der Folge $\{G_\nu\}$ werde zunächst angenommen, daß die vorgegebene torsionsfreie, abzählbare abelsche Gruppe B den Rang r hat. Gemäß Hilfssatz 1 sei B durch freie abelsche Untergruppen B_ν vom Rang r ausgeschöpft. Der Übergang von einer Basis $\mathfrak{t}_\nu$ in B_ν zu einer Basis $\mathfrak{t}_{\nu+1}$ in $B_{\nu+1}$ vollzieht sich vermöge der ganzzahligen, nicht singulären Matrix $\mathfrak{A}_{\nu+1}$:

$$\begin{aligned} t_{\nu 1} &= \lambda_{11}^{(\nu+1)} t_{\nu+1,1} \\ t_{\nu 2} &= \lambda_{21}^{(\nu+1)} t_{\nu+1,1} + \lambda_{22}^{(\nu+1)} t_{\nu+1,2} \\ &\vdots \\ t_{\nu r} &= \lambda_{r1}^{(\nu+1)} t_{\nu+1,1} + \cdots + \lambda_{rr}^{(\nu+1)} t_{\nu+1,r}. \end{aligned}$$

Wir legen $\{G_\nu\}$ induktiv fest.

Das Gebiet G_1 möge aus dem Dizylinder $D: \{|z_1| < 1, |z_2| < 1\}$ durch Herausnahme der analytischen Ebenenstücke $F_{1\varrho}: \left\{\varphi_{1\varrho}: = z_2 - \frac{r - \varrho}{2r} z_1 = 0, |z_1| < 1\right\}$, $\varrho = 1, \ldots, r$, entstehen. Es sei: $F_1: = \bigcup\limits_\varrho F_{1\varrho}$, $G_1: = D - F_1$, $\varphi_1: = \prod\limits_\varrho \varphi_{1\varrho}$. $U_1(F_1)$ sei eine gemäß Hilfssatz 2 gewählte Umgebung von F_1 in D, $\overline{U}_1$ ihre abgeschlossene Hülle in D. Wir setzen $G_1^*: = D - \overline{U}_1$.

Es werde angenommen, die Gebiete G_μ, $1 \leq \mu \leq \nu$, seien bereits bestimmt. G_ν habe folgende Eigenschaften:

(a) Es ist $G_\nu: = D - F_\nu$, $F_\nu: = \bigcup\limits_\varrho F_{\nu\varrho}$, $F_{\nu\varrho}$ ist der in D gelegene Teil der 1-dimensionalen analytischen Menge $\{\varphi_{\nu\varrho}(z_1, z_2) = 0\}$; dabei ist $\varphi_{\nu\varrho}$ ein Polynom in z_1, z_2, das die Voraussetzungen von Hilfssatz 2 erfüllt.

(b) Die $F_{\nu\varrho}$ sind irreduzibel in D.

(c) F_ν ist zusammenhängend.

Wir bestimmen wieder nach Hilfssatz 2 eine Umgebung $U_\nu = U_\nu(F_\nu)$ in D; es sei $G_\nu^* := D - \overline{U}_\nu$. Bei der Konstruktion von U_ν treten in der z_1-Ebene endlich viele punktfremde kritische Kreise $|z_1 - z_{\nu 1}^{(\lambda)}| \leq \delta(\varepsilon_1)$ auf. Ihr Komplement bezüglich $|z_1| < 1$ sei T_ν. Die Schnittpunkte von $E(z_1^{(\nu)}) : \{z_1 = z_1^{(\nu)}, |z_2| < 1\}$, $z_1^{(\nu)} \in T_\nu$, mit F_ν sind einfach. $E(z_1^{(\nu)}) \cap U_\nu(F_\nu)$ besteht aus endlich vielen offenen Kreisscheiben $V_{\nu\varkappa}$.

Bei Vorgabe dieser Daten werden $G_{\nu+1}$ und $G_{\nu+1}^*$ wie folgt konstruiert: $G_{\nu+1}$ soll wieder aus D durch Herausnahme von r in D irreduziblen 1-dimensionalen analytischen Mengen $F_{\nu+1,\varrho}$ entstehen, welche die Eigenschaften (a), (b), (c) haben; also $G_{\nu+1} := D - F_{\nu+1}, F_{\nu+1} := \bigcup_\varrho F_{\nu+1,\varrho}$. Überdies sollen noch folgende Eigenschaften erfüllt sein:

(d) $F_{\nu+1}$ besitzt eine $\varepsilon_{\nu+1}$-Umgebung $U_{\varepsilon_{\nu+1}}(F_{\nu+1})$ [vgl. (1) in Hilfssatz 2] mit $\overline{U}_{\varepsilon_{\nu+1}} \subset U_\nu(F_\nu)$.

(e) Ist $z_\varrho^{(\nu)} = (z_1^{(\nu)}, z_{\varrho 2}^{(\nu)}) \in F_{\nu\varrho} \cap E(z_1^{(\nu)})$, $z_1^{(\nu)} \in T_\nu$, und ist $V_{\nu\varrho}$ die Kreisscheibe von $E(z_1^{(\nu)}) \cap U_\nu(F_\nu)$ mit $z_\varrho^{(\nu)}$ als Mittelpunkt, $\varrho = 1, \ldots, r$, so ist die Schnittzahl[11]) $s(F_{\nu+1,\varrho}, V_{\nu\sigma}) = \lambda_{\sigma\varrho}^{(\nu+1)}$, $\sigma = 1, \ldots, r$ für jedes $\varrho = 1, \ldots, r$. Die nicht negativen ganzen Zahlen $\lambda_{\sigma\varrho}^{(\nu+1)}$ (vgl. Hilfssatz 1) sind für festes ϱ dabei genau die Elemente der ϱ-ten Spalte von $\mathfrak{A}_{\nu+1}$. Ist weiter $U_{\nu+1} = U_{\nu+1}(F_{\nu+1})$ eine gemäß Hilfssatz 2 konstruierte Umgebung, die in $U_{\varepsilon_{\nu+1}}(F_{\nu+1})$ enthalten ist, so ist $G_{\nu+1}^* := D - \overline{U}_{\nu+1}$. Die ε_ν, $\nu = 2, 3, \ldots$, sollen dabei eine monotone Nullfolge bilden.

Alles läuft also darauf hinaus, bei vorgegebenen 1-dimensionalen analytischen Mengen $F_{\nu\varrho}$ mit den Eigenschaften (a), (b), (c) neue 1-dimensionale analytische Mengen $F_{\nu+1,\varrho}$ mit den Eigenschaften (a), (b), (c), (d), (e) zu konstruieren. Wir werden diese Aufgabe in den letzten Abschnitten lösen (vgl. 6, 7, 8).

4. Nach Konstruktion bilden die $\overline{U}_\nu$ eine absteigende Folge von abgeschlossenen Teilmengen von D. Ihr Durchschnitt $F := \bigcap_\nu \overline{U}_\nu$ ist nicht leer und abgeschlossenen in D. Es sei $G := D - F$. Dann ist $G = \bigcup_\nu G_\nu^*$. Es sei weiter D_ν der offene Kern des Durchschnitts aller G_μ mit $\mu \geqq \nu$, $D_\nu := \overset{\circ}{\overbrace{\bigcap_{\mu \geqq \nu} G_\mu}}$. Dann ist $D_\nu = D - \overline{\bigcup_{\mu \geqq \nu} F_\mu}$. Es ist $D_\nu \subset D_{\nu+1}$, ferner $\bigcup_\nu D_\nu = G$. Beweis: Ist $z^{(0)} = (z_1^{(0)}, z_2^{(0)}) \in \bigcup_\nu D_\nu$, so gibt es ein ν_0, so daß $z^{(0)} \in D_{\nu_0} = D - \overline{\bigcup_{\mu \geqq \nu_0} F_\mu}$, also $z^{(0)} \notin \overline{\bigcup_{\mu \geqq \nu_0} F_\mu}$. Es seien $z_\varkappa^{(\mu)} = (z_1^{(0)}, z_{\varkappa 2}^{(\mu)})$, $\varkappa = 1, \ldots, k_\mu$, die Schnittpunkte von F_μ mit dem Ebenenstück $E(z_1^{(0)}) : \{z_1 = z_1^{(0)}, |z_2| < 1\}$, $|z_1^{(0)}| < 1$. Die $z_{\varkappa 2}^{(\mu)}$, $\mu \geqq \nu_0$, häufen sich nicht gegen $z_2^{(0)}$. Andrerseits werden die Kreisscheiben $V_{\mu\varkappa} := E(z_1^{(0)}) \cap \overline{U}_\mu$, die auf $E(z_1^{(0)})$ Umgebungen der $z_{\varkappa 2}^{(\mu)}$ bilden, nach Konstruktion

[11]) Die Schnittzahl zweier analytischer Flächen von komplementärer Dimension in bezug auf die durch die komplex-analytische Struktur induzierte Orientierung ist stets positiv, sofern $F_1 \cap F_2$ nicht leer ist. Vgl. [6], Fußnote 13.

der U_μ beliebig klein für μ hinreichend groß. Also gibt es ein $\nu_1 \geqq \nu_0$, so daß $z_2^{(0)} \notin E(z_1^{(0)}) \cap \overline{U}_{\nu_1}$, also $z_1^{(0)} \notin \overline{U}_{\nu_1}$, $z^{(0)} \notin \bigcap_\nu \overline{U}_\nu = F$, $z^{(0)} \in D - F$. Ist umgekehrt $z^{(0)} \in D - F$, so ist $z^{(0)} \notin \bigcap_\nu \overline{U}_\nu$. Es gibt also ein ν_0, so daß $z^{(0)} \notin \overline{U}_{\nu_0}$ und damit auch $z^{(0)} \notin \overline{U}_\mu$, $\mu \geq \nu_0$. Es ist $F_\mu \subset \overline{U}_\mu$ und $\overline{U}_{\mu+1} \subset \overline{U}_\mu$, also $\overline{\bigcup_{\mu \geq \nu_0} F_\mu} \subset \overline{U}_{\nu_0}$. Folglich ist $z^{(0)} \notin \overline{\bigcup_{\mu \geq \nu_0} F_\mu}$, also $z^{(0)} \in D - \overline{\bigcup_{\mu \geq \nu_0} F_\mu} = D_{\nu_0}$, $z^{(0)} \in \bigcup_\nu D_\nu$.

Zum Nachweis, daß G ein Holomorphiegebiet ist, benutzt man zweckmäßig die Darstellung $G = \bigcup_\nu D_\nu$ mit $D_\nu \subset D_{\nu+1}$. Wir zeigen, daß jedes D_ν Holomorphiegebiet ist. Es war $D_\nu := \overset{\circ}{\widehat{\bigcap_{\mu \geq \nu} G_\mu}} = D - \overline{\bigcup_{\mu \geq \nu} F_\mu}$ und $G_\mu := D - F_\mu$; dabei ist D das Holomorphiegebiet $\{|z_1| < 1, |z_2| < 1\}$ und F_μ der in D gelegene Teil der Nullstellenmenge des Polynoms $\varphi_\mu(z_1, z_2)$. Zunächst ist D_ν ein Gebiet. Dazu genügt es zu zeigen, daß der Durchschnitt von D_ν mit jedem Ebenenstück $E(z_1^{(0)}) : \{z_1 = z_1^{(0)}, |z_2| < 1\}$, $|z_1^{(0)}| < 1$, zusammenhängend ist. Es seien $z^{(0)} = (z_1^{(0)}, z_1^{(0)})$ und $\tilde{z}^{(0)} = (z_1^{(0)}, \tilde{z}_2^{(0)})$ aus $E(z_1^{(0)}) \cap D_\nu$. Analog wie beim Nachweis der Inklusion $\bigcup_\nu D_\nu \subset D - F$ sieht man, daß $z_2^{(0)} \notin E(z_1^{(0)}) \cap \overline{U}_{\nu_1}$ und $\tilde{z}_2^{(0)} \notin E(z_1^{(0)}) \cap \overline{U}_{\nu_2}$ mit geeigneten ν_1, ν_2. Sei etwa $\nu_1 \leqq \nu_2$, dann gilt wegen $\overline{U}_{\nu_2} \subset \overline{U}_{\nu_1}$: $z_2^{(0)}, \tilde{z}_2^{(0)} \notin E(z_1^{(0)}) \cap \overline{U}_{\nu_2}$. Dieser Durchschnitt besteht aus endlich vielen punktfremden Kreisscheiben, folglich ist sein Komplement bezüglich $E(z_1^{(0)})$ zusammenhängend.

Die G_μ sind Holomorphiegebiete, folglich auch D_ν als Durchschnitt von Holomorphiegebieten[12]), dann ist aber nach einem bekannten Satz von BEHNKE-STEIN[13]) auch G ein Holomorphiegebiet.

Zum Nachweis, daß G die gewünschte topologische Eigenschaft hat, benutzen wir die Gleichung $G = \bigcup_\nu G_\nu^*$ mit $G_\nu^* \subset G_{\nu+1}^*$. Es war $U_\nu = U_\nu(F_\nu)$ eine Umgebung von F_ν mit $U_\nu \supset U_{\nu+1}$ und $G_\nu^* := D - \overline{U}_\nu$. Wir setzen noch $\tilde{G}_\nu := D - U_\nu$. Es ist $\tilde{G}_\nu \subset G_{\nu+1}^* \subset G$, wegen $\overline{U}_{\nu+1} \subset \overline{U}_{\varepsilon_{\nu+1}} \subset U_\nu$. Die 1. Bettische Gruppe $B_1(G_\nu)$ von $G_\nu := D - F_\nu$ ist nach Satz 2 ein r-dimensionales Gitter. Die 1-dimensionalen Zyklen $\alpha_{\nu 1}^1, \ldots, \alpha_{\nu r}^1$ seien Repräsentanten einer Homologiebasis von $B_1(G_\nu)$. $\alpha_{\nu\varrho}^1$ bezeichne nach Bedarf den Zyklus und seine Homologieklasse. Ist $z_\varrho^{(\nu)} = (z_1^{(\nu)}, z_{\varrho 2}^{(\nu)}) \in F_{\nu\varrho} \cap E(z_1^{(\nu)})$ mit $E(z_1^{(\nu)}) : \{z_1 = z_1^{(\nu)}, |z_2| < 1\}$, $z_1^{(\nu)} \in T_\nu$, und $V_{\nu\varrho}$ die Kreisscheibe von $E(z_1^{(\nu)}) \cap U_\nu(F_\nu)$ mit $z_\varrho^{(\nu)}$ als Mittelpunkt, $\varrho = 1, \ldots, r$, so kann man den positiv orientierten Rand der $V_{\nu\varrho}$ als Träger der $\alpha_{\nu\varrho}^1$ wählen. Die $\alpha_{\nu\varrho}^1$ bilden nach Hilfssatz 2 gleichzeitig ein freies Erzeugendensystem von $B_1(\tilde{G}_\nu)$. Wir behaupten, daß sie für $\nu = 1, 2, 3, \ldots; \varrho = 1, \ldots, r$ auch ein Erzeugendensystem von $B_1(G)$ bilden. Zunächst ist kein ganzzahliges Vielfaches $m (m \neq 0)$ eines $\alpha_{\nu\varrho}^1$ in G nullhomolog. Wäre etwa $m\alpha_{\nu\varrho}^1$ Rand einer 2-Kette β^2 in G, so wäre der kompakte Träger von β^2 bereits in einem $G_{\nu_0}^*$, $\nu_0 > \nu$ hinreichend groß, enthalten und damit auch in $\tilde{G}_{\nu_0}$. Dann wäre bereits in

---

12) Vgl. [3], p. 74.

13) Ist H ein Gebiet im C^2 und $\{H_\nu\}$ mit $H_\nu \subset H_{\nu+1}$ eine Ausschöpfungsfolge von H durch Holomorphiegebiete H_ν, so ist auch H ein Holomorphiegebiet. Vgl. hierzu [2].

$\tilde{G}_{\nu_k}$ $m\alpha^1_{\nu\varrho} \sim 0$. Andrerseits ist $\alpha^1_{\nu\varrho}$ als Zyklus in $\tilde{G}_{\nu_k}$ einer ganzzahligen Linearkombination der $\alpha^1_{\nu,\varrho}$ homolog. Das ist im Widerspruch zur homologen Unabhängigkeit der $\alpha^1_{\nu,\varrho}$ in $\tilde{G}_{\nu}$. Ist weiter α^1 irgendein 1-Zyklus in G, so ist sein kompakter Träger in einem geeigneten $G^*_{\nu_k}$ enthalten und damit auch in $\tilde{G}_{\nu_k}$. In G_{ν_k} und damit auch in $G \supset \tilde{G}_{\nu_k}$ ist α^1 einer ganzzahligen Linearkombination der $\alpha^1_{\nu,\varrho}$, $\varrho = 1, \ldots, r$, homolog. Die Zyklen $\alpha^1_{\nu\varrho}$ sind auch Zyklen in $\tilde{G}_{\nu+1}$. Wegen Eigenschaft (e) der $F_{\nu+1,\varrho}$ bestimmt die Matrix $\mathfrak{A}_{\nu+1}$ genau die Homologierelationen zwischen den $\alpha^1_{\nu\varrho}$ und den $\alpha^1_{\nu+1,\varrho}$ in $\tilde{G}_{\nu+1}$ und damit auch in G. Man wähle dazu $z_1^{(0)} \in T_\nu \cap T_{\nu+1}$ und betrachte den Schnitt des Ebenenstücks $E(z_1^{(0)}) : \{z_1 = z_1^{(0)}, |z_2| < 1\}$ mit $\tilde{G}_\nu$ und $\tilde{G}_{\nu+1}$.

5. Hat die vorgegebene abzählbare, torsionsfreie abelsche Gruppe B unendlichen Rang, so läßt sich die Konstruktion von G wie folgt durchführen. Man schöpfe B durch freie abelsche Gruppen B_ν vom Rang r_ν aus. Wie beim Beweis von Hilfssatz 1 gehe man etwa von r linear unabhängigen Elementen aus. Die von ihnen erzeugte Untergruppe sei B_1. Es ist $r_1 = r$. Sind die B_μ mit $B_{\mu-1} \subset B_\mu$ für $\mu \leq \nu$ bereits bestimmt, so streiche man in einer mit den Basiselementen von B_1 beginnenden Abzählung I von B sämtliche Elemente von B_ν. Sei t das erste nicht gestrichene Element und $B_{\nu+1}$ die von den Basiselementen $t_{\nu 1}, \ldots, t_{\nu r_\nu}$ von B_ν und von t erzeugte Untergruppe von B. Sind t und $t_{\nu 1}, \ldots, t_{\nu r_\nu}$ linear abhängig, dann ist $r_{\nu+1} = r_\nu$, und der Zusammenhang zwischen einer Basis von B_ν und einer Basis von $B_{\nu+1}$ kann nach Hilfssatz 1 durch eine Dreiecksmatrix mit positiven Elementen in und unterhalb der Hauptdiagonalen beschrieben werden. Sind dagegen $t_{\nu 1}, \ldots, t_{\nu r_\nu}$, t linear unabhängig, so ist $r_{\nu+1} = r_\nu + 1$. Als Basis von $B_{\nu+1}$ wählen wir die Elemente $t_{\nu+1,\varrho} := t_{\nu\varrho}$ für $\varrho = 1, \ldots, r_\nu - 1$; $t_{\nu+1,r_\nu} := t_{\nu r_\nu} - t$; $t_{\nu+1,r_{\nu+1}} := t$.

Entsprechend sind bei der Konstruktion von $G_{\nu+1}$ bei gegebenem G_ν (G_1 ist wie früher gewählt) zwei Fälle zu unterscheiden. G_ν erfülle wieder die Eigenschaften (a), (b), (c), und es sei $G^*_\nu := D - \overline{U}_\nu$, wobei $U_\nu = U_\nu(F_\nu)$ nach Hilfssatz 2 bestimmt ist. Ist dann $r_{\nu+1} = r_\nu$, so ändert sich an der bisherigen Konstruktion von $G_{\nu+1}$ gar nichts. Ist $r_{\nu+1} = r_\nu + 1$, so sind $r_{\nu+1}$ in D irreduzible 1-dimensionale analytische Mengen $F_{\nu+1,\varrho}$, $\varrho = 1, \ldots, r_{\nu+1}$, gesucht, so daß mit $F_{\nu+1} := \bigcup\limits_{\varrho} F_{\nu+1,\varrho}$ und $G_{\nu+1} := D - F_{\nu+1}$ die Eigenschaften (a) bis (d) erfüllt sind. Ferner soll gelten:

(e′) Ist $z_\varrho^{(\nu)} = (z_1^{(\nu)}, z_{\varrho 2}^{(\nu)}) \in F_{\nu\varrho} \cap E(z_1^{(\nu)})$, $z_1^{(\nu)} \in T_\nu$, und ist $V_{\nu\varrho}$ die Kreisscheibe von $E(z_1^{(\nu)}) \cap U_\nu(F_\nu)$ mit $z_\varrho^{(\nu)}$ als Mittelpunkt, $\varrho = 1, \ldots, r_\nu$, so ist:

$$s(F_{\nu+1,\varrho}, V_{\nu\sigma}) = \begin{cases} 1 \text{ für } \sigma = \varrho \\ 0 \text{ für } \sigma \neq \varrho, \sigma = 1, \ldots, r_\nu, \varrho = 1, \ldots, r_\nu . \end{cases}$$

$$s(F_{\nu+1,r_{\nu+1}}, V_{\nu\sigma}) = \begin{cases} 1 \text{ für } \sigma = r_\nu \\ 0 \text{ für } \sigma = 1, \ldots, r_\nu - 1 . \end{cases}$$

$U_{\nu+1} = U_{\nu+1}(F_{\nu+1})$ wird nach Hilfssatz 2 wieder so bestimmt, daß $U_{\nu+1} \subset U_{\varepsilon_{\nu+1}}$ gilt. Es ist $G^*_{\nu+1} := D - \overline{U}_{\nu+1}$. Die ε_ν, $\nu = 2, 3, \ldots$, sollen eine monotone Nullfolge bilden.

Wir setzen wie in Abschnitt 4 $G := \bigcup_{\nu} G_\nu^*$ mit $G_\nu^* \subset G_{\nu+1}^*$. Der Nachweis, daß G Holomorphiegebiet ist, bleibt ungeändert. Auch am Nachweis der Isomorphie $B_1(G) \cong B$ ändert sich nichts. Das System der genau wie dort gewählten 1-Zyklen $\alpha_{\nu\varrho}^1$, $\nu = 1, 2, \ldots$; $\varrho = 1, \ldots, r_\nu$, bildet ein Erzeugendensystem von $B_1(G)$. Zu erwähnen ist noch der Fall $r_{\nu+1} = r_\nu + 1$. Wegen (e′) gilt in $\widetilde{G}_{\nu+1}$ und damit in G: $\alpha_{\nu\varrho}^1 \sim \alpha_{\nu+1,\varrho}^1$ für $\varrho = 1, \ldots, r_\nu - 1$; $\alpha_{\nu r_\nu}^1 \sim \alpha_{\nu+1, r_\nu}^1 + \alpha_{\nu+1, r_\nu+1}^1$. Das ist aber für diesen Fall genau der Zusammenhang zwischen den Basiselementen von B_ν und $B_{\nu+1}$.

6. Zum endgültigen Beweis von Satz 3 bleibt noch die Konstruktion 1-dimensionaler analytischer Mengen $F_{\nu+1,\varrho}$ mit den Eigenschaften (a) bis (e) bzw. (e′) bei gegebenen 1-dimensionalen analytischen Mengen $F_{\nu\varrho}$ mit den Eigenschaften (a), (b), (c). Wir nehmen zunächst wieder an: Rang $(B) = r$. Der Fall $\varrho = r$ ist gesondert zu behandeln; wir geben die $F_{\nu r}$, $\nu = 1, 2, \ldots$, direkt an. Es ist F_{1r} die Ebene $z_2 = 0$. Ferner gilt nach (e):

$$s(F_{\nu+1,r}, V_{\nu\sigma}) = \begin{cases} 0 \text{ für } \sigma = 1, \ldots, r-1 \\ \lambda_{rr}^{(\nu+1)} \text{ für } \sigma = r\,. \end{cases}$$

Wir setzen $\lambda_{\nu+1} := \lambda_{rr}^{(\nu+1)}$ und bilden mit komplexen $b_\nu \neq 0$

$$z_2 = f_\nu(z_1) := b_2 z_1^{\frac{1}{\lambda_2}} + \cdots + b_\nu z_1^{\frac{1}{\lambda_2 \cdot \ldots \cdot \lambda_\nu}} \text{ für } \nu \geqq 2.$$

Dabei sei stets

$$z_1^{\frac{1}{\lambda_2 \cdot \ldots \cdot \lambda_{\nu-1}}} = \left(z_1^{\frac{1}{\lambda_2 \cdot \ldots \cdot \lambda_\nu}}\right)^{\lambda_\nu} \text{ für } \nu > 2.$$

Die so definierte algebraische Funktion genügt einer irreduziblen algebraischen Gleichung $\varphi_{\nu r}(z_1, z_2) = 0$. Der Koeffizient der höchsten Potenz in z_2 ist dabei 1. $F_{\nu r}$ sei der in D gelegene Teil der durch $\varphi_{\nu r} = 0$ bestimmten analytischen Menge.

Bei der Konstruktion der $F_{\nu+1,\varrho}$, $\varrho = 1, \ldots, r-1$, sind zwei Fälle zu unterscheiden, je nachdem ob die $\lambda_{\varrho\varrho}^{(\nu+1)}, \ldots, \lambda_{r\varrho}^{(\nu+1)}$ teilerfremd sind oder nicht. Wir erledigen zunächst den Fall, daß die $\lambda_{\varrho\varrho}^{(\nu+1)}, \ldots, \lambda_{r\varrho}^{(\nu+1)}$, $\varrho = 1, \ldots, r-1$, keinen gemeinsamen Teiler haben. Den Index $(\nu+1)$ lassen wir weg, wenn keine Verwechslung eintreten kann.

$$\varphi_{\nu 1}(z_1, z_2) = 0, \ldots, \varphi_{\nu r}(z_1, z_2) = 0$$

seien Darstellungen der $F_{\nu 1}, \ldots, F_{\nu r}$. Dann wird $F_{\nu+1,\varrho}$, $\varrho = 1, \ldots, r-1$ durch folgende Gleichung bestimmt:

$$(*) \qquad \varphi_{\nu+1,\varrho}(z_1, z_2) := \varphi_{\nu\varrho}^{\lambda_{\varrho\varrho}}(z_1, z_2) \cdot \varphi_{\nu,\varrho+1}^{\lambda_{\varrho+1,\varrho}}(z_1, z_2) \cdot \ldots \cdot \varphi_{\nu r}^{\lambda_{r\varrho}}(z_1, z_2) - \delta_{\nu+1,\varrho} = 0$$

$\delta_{\nu+1,\varrho} \neq 0$, komplex.

$F_{\nu+1}$ ist der in D gelegene Teil der Nullstellenmenge des Polynoms $\varphi_{\nu+1}(z_1, z_2) := \prod_{\varrho=1}^{r} \varphi_{\nu+1,\varrho}(z_1, z_2)$. $U_\nu(F_\nu)$ enthält eine ε-Umgebung $U_\varepsilon(F_\nu)$. Wir wählen die $\delta_{\nu+1,1}, \ldots, \delta_{\nu+1,r-1}$ und $b_{\nu+1}$ so klein, daß $F_{\nu+1}$ in $U_{\varepsilon/2}(F_\nu)$ enthalten ist. Dann gibt es stets eine Umgebung $\overline{U}_{\varepsilon_{\nu+1}}(F_{\nu+1})$, so daß Eigenschaft (d) erfüllt ist.

Daß die Nullstellenmengen der Polynome $\varphi_{\nu+1,\varrho}(z_1, z_2)$, $\varrho = 1, \ldots, r$, mit $S: \{|z_1| \leqq 1, |z_2| = 1\}$ einen leeren Durchschnitt haben, ist auf Grund der

Wahl von $\varphi_1(z_1, z_2)$ und wegen $U_{\nu+1} \subset U_\nu$ klar. Ferner sind die $\varphi_{\nu+1,\varrho}(z_1, z_2)$ Polynome der Bauart $z_2^n + a_{n-1}(z_1)\, z_2^{n-1} + \ldots + a_0(z_1)$, $a_\nu(z_1)$ Polynome in z_1. Für $\varphi_{\nu+1,r}$ ist es bereits bewiesen, für die übrigen $\varphi_{\nu+1,\varrho}$, $\varrho = 1, \ldots, r-1$, folgt es aus (*). Also ist Eigenschaft (a) erfüllt. Eigenschaft (c) folgt unmittelbar aus (*). Eigenschaft (b) ergibt sich aus dem folgenden

Hilfssatz 3: *Im Dizylinder* $D: \{|z_1| < 1, |z_2| < 1\}$ *seien* $m > 1$ *irreduzible analytische Mengen* A_μ *durch Gleichungen*

$$\psi_\mu(z_1, z_2) = 0, \quad \mu = 1, \ldots, m$$

gegeben; dabei seien die ψ_μ *in* D *holomorphe Funktionen.* $A := \bigcup_{\mu=1}^{m} A_\mu$ *sei zusammenhängend. Sind dann* $k_1, \ldots, k_m$ *teilerfremde natürliche Zahlen, so gibt es ein* $\varepsilon > 0$*, so daß für jede komplexe Zahl* $t \neq 0$ *mit* $|t| < \varepsilon$ *die durch die Nullstellenmenge der Funktion*

$$\psi_t(z_1, z_2) := \left(\prod_{\mu=1}^{m} \psi_\mu^{k_\mu}(z_1, z_2)\right) - t$$

in D *definierte analytische Menge* $A(t)$ *irreduzibel in* D *ist.*

Beweis: Es sei $z^{(\mu)}$ ein gewöhnlicher Punkt von A_μ, $\mu = 1, \ldots, m$. Durch jeden dieser Punkte legen wir eine 1-dimensionale analytische Ebene E_μ, die jeweils in einer Umgebung U_μ von $z^{(\mu)}$ nur den Punkt $z^{(\mu)}$ mit A_μ gemeinsam hat. Die U_μ seien punktfremd gewählt. Für hinreichend kleine t ($|t| < \varepsilon$) schneidet auch $A(t)$ die Ebene E_μ in U_μ jeweils nur in isolierten Punkten, und zwar ist in U_μ die Schnittzahl $s(A(t), E_\mu) = k_\mu$. Nach einem von K. Stein bewiesenen Approximationssatz[14]) können wir $\varepsilon > 0$ von vornherein so klein wählen, daß jede irreduzible Komponente von $A(t)$ für t aus $|t| < \varepsilon$ mit jeder der Ebenen E_μ in U_μ gemeinsame Punkte hat. Wir nehmen jetzt an, $A(t)$ zerfalle für irgendein solches $t \neq 0$ in die irreduziblen Komponenten $A_\pi(t)$, $\pi = 1, \ldots, p$. Dann ist $l_\pi^{(\mu)} := s(A_\pi(t), E_\mu) > 0$ und $\sum_{\pi=1}^{p} l_\pi^{(\mu)} = k_\mu$. Da $A_\pi(t)$ und $A_{\pi'}(t)$ stetig ineinander deformierbar sind vermöge einer Schar, deren Elemente stets die Schnittzahlbedingung erfüllen, so folgt: $l_\pi^{(\mu)} = l_{\pi'}^{(\mu)}$[15]). Folglich gilt also, wenn man setzt: $l_\mu := l_\pi^{(\mu)} = l_{\pi'}^{(\mu)}$, $k_\mu = p \cdot l_\mu$ mit $p > 1$. Das ist ein Widerspruch zur Teilerfremdheit der k_μ, also muß $A(t)$ für $t \neq 0$ aus $|t| < \varepsilon$ irreduzibel sein.

---

[14]) Der Satz lautet (vgl. [13]): *In der zusammenhängenden n-dimensionalen komplexen Mannigfaltigkeit X sei eine zusammenhängende rein k-dimensionale analytische Menge A vorgegeben, es sei* $k < n$. *Je zwei irreduzible Komponenten von A mögen sich in X in unabhängiger Lage befinden. In X seien weiter rein k-dimensionale analytische Mengen* A_j $(j \in J)$ *gegeben, die zu A punktfremd sind. Die* A_j *mögen keinen Häufungspunkt in* $X - A$*, wohl aber einen Konvergenzpunkt auf A besitzen. Dann konvergiert die Folge der* A_j *gegen A.*

Dabei heißen zwei irreduzible, rein k-dimensionale analytische Mengen A und $'A$ in X in *unabhängiger Lage*, wenn in jedem Punkt $x \in A \cap {'A}$ gilt: $\dim(A \cap {'A}) = \dim A + \dim {'A} - n$. (Zur genauen Def. der Dimension einer analytischen Menge siehe [8].)

Zwei rein $(n-1)$-dimensionale analytische Mengen in X befinden sich stets in unabhängiger Lage, ebenso zwei punktfremde analytische Mengen in X.

[15]) Vgl. [6], Fußnote 13.

Um Eigenschaft (c) einzusehen, betrachte man den Schnitt eines Ebenenstücks $E(z_1^{(\nu)}) : \{z_1 = z_1^{(\nu)}, |z_2| < 1\}$, $z_1^{(\nu)} \in T_\nu$, mit $F_{\nu r}$. Es sei $z_r^{(\nu)} = (z_1^{(\nu)}, z_{r2}^{(\nu)}) \in F_{\nu r} \cap E(z_1^{(\nu)})$ und $V_{\nu r}$ die Kreisscheibe von $U_\nu(F_\nu) \cap E(z_1^{(\nu)})$, die $z_r^{(\nu)}$ als Mittelpunkt hat. Es ist nach (e) $s(F_{\nu+1,\varrho}, V_{\nu r}) = \lambda_{r\varrho}^{(\nu+1)} > 0$ für $\varrho = 1, \ldots, r$. Die komplexen von null verschiedenen Zahlen $\delta_{\nu+1,\varrho}, b_{\nu+1}, \varrho = 1, \ldots, r-1$, seien bereits so gewählt, daß die Eigenschaften (a), (b), (d), (e) auch für alle von null verschiedenen Zahlen $\delta^*_{\nu+1,\varrho}, b^*_{\nu+1}$ mit $|\delta^*_{\nu+1,\varrho}| < |\delta_{\nu+1,\varrho}|$ und $|b^*_{\nu+1}| < |b_{\nu+1}|$ erfüllt sind. Das ist nach Hilfssatz 3 möglich. Eine eventuelle weitere Verkleinerung der $\delta_{\nu+1,\varrho}, b_{\nu+1}$ ist also erlaubt. Mit ihrer Hilfe läßt sich stets erreichen, daß sämtliche $F_{\nu+1,\varrho}, \varrho = 1, \ldots, r$, einen gemeinsamen Punkt auf $V_{\nu r}$ haben.

7. Haben die Elemente der $r-1$ ersten Spalten von $\mathfrak{A}_{\nu+1}$ jeweils die größten gemeinsamen Teiler $d_1^{(\nu+1)}, \ldots, d_{r-1}^{(\nu+1)}$, so setze man

$$\check{\lambda}_{\sigma\varrho}^{(\nu+1)} := \frac{1}{d_\varrho^{(\nu+1)}} \lambda_{\sigma\varrho}^{(\nu+1)}, \sigma = 1, \ldots, r; \varrho = 1, \ldots, r-1, \check{\lambda}_{\sigma r}^{(\nu+1)} := \lambda_{\sigma r}^{(\nu+1)}.$$

$$\check{t}_{\nu+1,\lambda} := d_\varrho^{(\nu+1)} t_{\nu+1,\varrho}, \check{t}_{\nu+1,r} := t_{\nu+1,r}.$$

Wenigstens eine der natürlichen Zahlen $d_\varrho^{(\nu+1)}$ sei größer als 1. Die $\check{t}_{\nu+1,\varrho}$, $\varrho = 1, \ldots, r$, sind Basiselemente eines Gitters $\check{B}_{\nu+1}$ mit $B_\nu \subset \check{B}_{\nu+1} \subset B_{\nu+1}$. Es ist

$$t_{\nu\varrho} = \sum_{\sigma=1}^{r} \check{\lambda}_{\varrho\sigma}^{(\nu+1)} \check{t}_{\nu+1,\sigma}, \varrho = 1, \ldots, r,$$

oder in Matrizenschreibweise: $t_\nu = \check{\mathfrak{A}}_{\nu+1} \check{t}_{\nu+1}$.

Die Elemente der Spalten von $\check{\mathfrak{A}}_{\nu+1}$ sind teilerfremd. Folglich lassen sich auf die in Abschnitt 6 beschriebene Art analytische Mengen $\check{F}_{\nu+1,\varrho}, \varrho = 1, \ldots, r$, konstruieren und damit auch die Gebiete $\check{G}_{\nu+1} := D - \check{F}_{\nu+1}$ und $\check{G}^*_{\nu+1} := D - \overline{U}_{\nu+1}(\check{F}_{\nu+1})$. Ferner ist

$$\check{t}_{\nu+1,\varrho} = d_\varrho^{(\nu+1)} t_{\nu+1,\varrho}, \varrho = 1, \ldots, r; d_r^{(\nu+1)} = 1.$$

Gesucht sind bei gegebenen analytischen Mengen $\check{F}_{\nu+1,\varrho}$ mit den Eigenschaften (a) bis (e) neue analytische Mengen $F_{\nu+1,\varrho}$ mit den Eigenschaften (a) bis (d) und der folgenden Eigenschaft

(e'') Ist $\check{z}_\varrho^{(\nu+1)} = (\check{z}_1^{(\nu+1)}, \check{z}_{\varrho 2}^{(\nu+1)}) \in \check{F}_{\nu+1,\varrho} \cap E(\check{z}_1^{(\nu+1)})$, $\check{z}_1^{(\nu+1)} \in \check{T}_{\nu+1}$, und ist $\check{V}_{\nu+1,\varrho}$ die Kreisscheibe von $E(\check{z}_1^{(\nu+1)}) \cap U_{\nu+1}(F_{\nu+1})$ mit $\check{z}_\varrho^{(\nu+1)}$ als Mittelpunkt, $\varrho = 1, \ldots, r$, so ist

$$s(F_{\nu+1,\varrho}, \check{V}_{\nu+1,\sigma}) = \begin{cases} d_\varrho^{(\nu+1)} & \text{für } \sigma = \varrho \\ 0 & \text{für } \sigma \neq \varrho. \end{cases}$$

Die $\check{F}_{\nu+1,\varrho}$ sind gegeben durch irreduzible algebraische Gleichungen $\check{\varphi}_{\nu+1,\varrho}(z_1, z_2) = 0$.

Ist $d_\varrho^{(\nu+1)} = 1$, so setze man $F_{\nu+1,\varrho} := \check{F}_{\nu+1,\varrho}$.

Ist $d_\varrho^{(\nu+1)} > 1$, ferner $z_2 = \check{f}_{\nu+1,\varrho}(z_1)$ die zu $\check{\varphi}_{\nu+1,\varrho} = 0$ gehörige algebraische Funktion, so bilde man die algebraische Funktion

$$z_2 = f_{\nu+1,\varrho}(z_1) := \check{f}_{\nu+1,\varrho}(z_1) + b_{\nu+1,\varrho} \cdot (z_1 - z_1^{(0)})^{\frac{1}{d_\varrho^{(\nu+1)}}}$$

mit komplexem $b_{\nu+1,\varrho} \neq 0$. $z_1^{(0)}$ mit $|z_1^{(0)}| < 1$ sei dabei kein Verzweigungspunkt von $\check{f}_{\nu+1,\varrho}(z_1)$. Es sei $\varphi_{\nu+1,\varrho}(z_1, z_2) = 0$ die zu $f_{\nu+1,\varrho}(z_1)$ gehörige irreduzible algebraische Gleichung. Der Koeffizient der höchsten Potenz in z_2 kann wieder zu 1 normiert werden, da $f_{\nu+1,\varrho}(z_1)$ über keinem endlichen z_1 einen Pol hat. $F_{\nu+1,\varrho}$ sei der in D gelegene Teil der Nullstellenmenge von $\varphi_{\nu+1,\varrho}(z_1, z_2)$. Wird $b_{\nu+1,\varrho} \neq 0$ hinreichend klein gewählt, so erfüllt $F_{\nu+1} := \bigcup_\varrho F_{\nu+1,\varrho}$ die Eigenschaften (a), (b), (d), (e″). Auch der Zusammenhang [Eigenschaft (c)] läßt sich durch geeignete Wahl der $b_{\nu+1,\varrho}$ erreichen. Damit sind auch für diesen Fall die Gebiete $G_{\nu+1} := D - F_{\nu+1}$ und $G^*_{\nu+1} := D - \overline{U}_{\nu+1}(F_{\nu+1})$ mit $\check{G}^*_{\nu+1} \subset G^*_{\nu+1}$ bestimmt.

8. Es bleibt der Fall, daß die abelsche Gruppe B unendlichen Rang hat (vgl. Abschnitt 5). Ist $r_{\nu+1} = r_\nu$, so konstruiere man $F_{\nu+1}$ wie in den vorhergehenden Abschnitten 6 bzw. 7 angegeben. Ist $r_{\nu+1} = r_\nu + 1$, so sind $r_{\nu+1}$ analytische Mengen $F_{\nu+1,\varrho}$, $\varrho = 1, \ldots, r_{\nu+1}$, gesucht, die neben den Eigenschaften (a) bis (d) noch (e′) erfüllen. Wir setzen $F_{\nu+1,\varrho} := F_{\nu\varrho}$ für $\varrho = 1, \ldots, r_\nu$. $F_{\nu r_\nu}$ ist der in D gelegene Teil der Nullstellenmenge des Polynoms $\varphi_{\nu r_\nu}(z_1, z_2)$. Wir bilden $\varphi_{\nu+1, r_{\nu+1}} := \varphi_{\nu r_\nu} - \delta_{\nu+1, r_{\nu+1}}$, $\delta_{\nu+1, r_{\nu+1}} \neq 0$ komplex. $F_{\nu+1, r_{\nu+1}}$ sei der in D gelegene Teil einer irreduziblen Komponente der Nullstellenmenge von $\varphi_{\nu+1, r_{\nu+1}}$. Bei geeigneter, hinreichend kleiner Wahl von $\delta_{\nu+1, r_{\nu+1}}$ erfüllt $F_{\nu+1} := \bigcup_{\varrho=1}^{r_{\nu+1}} F_{\nu+1,\varrho}$ die Eigenschaften (a) bis (e′).

Literatur

[1] Alexandroff, P., u. H. Hopf: Topologie I. Berlin: Springer 1935. — [2] Behnke, H., u. K. Stein: Konvergente Folgen von Regularitätsbereichen und die Meromorphiekonvexität. Math. Ann. **116**, 204—216 (1939). — [3] Behnke, H., u. P. Thullen: Theorie der Funktionen mehrerer komplexer Veränderlichen. Erg. d. Math. 3. Berlin: Springer 1934. — [4] Cartan, H.: Séminaire E. N. S. Paris 1948/49 (hektographiert). — [5] Dolbeault, P.: Formes différentielles et cohomologie sur une variété analytique complexe, I. Ann. of Math. **64**, no. 1, 83—130 (1956). — [6] Grauert, H., u. R. Remmert: Konvexität in der komplexen Analysis. Comm. Math. Helv. **31**, 152—183 (1956). — [7] Pontrjagin, L.: Über den algebraischen Inhalt topologischer Dualitätssätze. Math. Ann. **105**, 165—205 (1931). — [8] Remmert, R., u. K. Stein: Über die wesentlichen Singularitäten analytischer Mengen. Math. Ann. **126**, 263—306 (1953). — [9] Remmert, R.: Projektionen analytischer Mengen. Math. Ann. **130**, 410—441 (1956). — [10] Seifert, H., u. W. Threlfall: Lehrbuch der Topologie. Leipzig u. Berlin: C. G. Teubner 1934. — [11] Serre, J. P.: Quelques problèmes globaux relatifs aux variétés de Stein. Coll. de Bruxelles, 57—68 (1953). — [12] Stein, K.: Analytische Funktionen mehrerer komplexer Veränderlichen zu vorgegebenen Periodizitätsmoduln und das zweite Cousinsche Problem. Math. Ann. **123**, 201—222 (1951). — [13] Stein, K.: Ein Konvergenzsatz für analytische Mengen. In Vorbereitung.

(Eingegangen am 2. Juni 1959)

RIEGER, G. J.
Math. Annalen 138, 356—362 (1959)

Über Partitionen

Von

G. J. RIEGER in College Park, Maryland

Wir bezeichnen mit $p_k(n)$ die Anzahl aller Darstellungen einer natürlichen Zahl n als Summe von k natürlichen Zahlen, wobei alle Darstellungen, die aus einer bestimmten durch alleinige Änderung der Anordnung entstehen, mit dieser als identisch betrachtet werden. Mit anderen Worten: $p_k(n)$ ist die Lösungsanzahl der Diophantischen Gleichung

$$(0.1)\qquad n_1 + n_2 + \cdots + n_k = n \qquad (1 \leqq n_1 \leqq n_2 \leqq \cdots \leqq n_k)\,.$$

So ist z. B. $p_3(5) = 2$, $p_3(6) = 3$. Ferner ist $p_1(n) = 1$ für jedes n und $p_k(n) = 0$ für jedes $n < k$.

ISEKI hat 1940

$$(0.2)\qquad p_k(n) = \frac{n^{k-1}}{k!\,(k-1)!}\,(1 + o(1)) \qquad \begin{pmatrix} k \text{ fest} \\ n \to \infty \end{pmatrix}$$

bewiesen[1]), was 1942 von GUPTA zu

$$(0.3)\qquad \frac{1}{k!}\binom{n-1}{k-1} \leqq p_k(n) \leqq \frac{1}{k!}\binom{n + \binom{k}{2}}{k-1}$$

verschärft wurde[2]). Dieser letzteren Aussage entspricht die Entwicklung[3])

$$(0.4)\qquad \frac{n^{k-1}}{k!\,(k-1)!}\left(1 - \frac{k(k-1)}{2n} + O(n^{-2})\right) \leqq$$

$$\leqq p_k(n) \leqq \frac{n^{k-1}}{k!\,(k-1)!}\left(1 + \frac{(k-1)(k^2-2k+2)}{2n} + O(n^{-2})\right) \qquad \begin{pmatrix} k \geqq 3, \text{ fest} \\ n \to \infty \end{pmatrix}.$$

Das Ziel dieser Note ist es, einige weiterführende Sätze über $p_k(n)$ zu beweisen.

1. Zunächst gilt in Verallgemeinerung von (0.2) und (0.4)

Satz 1. *Es ist*

$$p_k(n) = \frac{n^{k-1}}{k!\,(k-1)!}\left(1 + \frac{k(k-1)(k-3)}{4n} + O(n^{-2})\right) \qquad \begin{pmatrix} k \geqq 3, \text{ fest} \\ n \to \infty \end{pmatrix}.$$

Hilfssatz 1. *Es ist*

$$p_k(n) = p_{k-1}(n-1) + p_k(n-k) = \sum_{0 \leqq k\mu \leqq n-k} p_{k-1}(n-1-k\mu)\,.$$

[1]) Vgl. [1], 52.
[2]) Vgl. [1], 52.
[3]) Die Konstante in $O(..)$ hängt höchstens von k ab.

Beweis. Die Anzahl aller Darstellungen (0.1), wo mindestens ein $n_\varkappa = 1$ ist, ist $p_{k-1}(n-1)$; die Anzahl aller Darstellungen (0.1), wo alle $n_\varkappa > 1$ sind, ist $p_k(n-k)$. Das beweist den ersten Teil; der zweite Teil folgt daraus unmittelbar.

Beweis von Satz 1. Wegen

$$p_2(n) = \left[\frac{n}{2}\right] \tag{1.1}$$

ergibt sich aus Hilfssatz 1

$$p_3(n) = \sum_{0 \leq 3\mu \leq n-3} p_2(n-1-3\mu) = \left[\frac{n^2+3}{12}\right]. \tag{1.2}$$

Für $k = 3$ ist Satz 1 also richtig, und wir nehmen jetzt an, er sei schon für $k-1$ bewiesen:

$$\begin{aligned} &p_{k-1}(n) = \\ &= \frac{1}{(k-1)!\,(k-2)!}\left(n^{k-2} + \frac{(k-1)(k-2)(k-4)}{4}\, n^{k-3} + O(n^{k-4})\right) \qquad \begin{pmatrix} k \geq 4, \text{ fest} \\ n \to \infty \end{pmatrix}. \end{aligned} \tag{1.3}$$

Nach Hilfssatz 1 ist dann

$$\begin{aligned} p_k(n) = \frac{1}{(k-1)!\,(k-2)!} \sum_{0 \leq k\mu \leq n-k} \Big((n-1-k\mu)^{k-2} + \\ + \frac{(k-1)(k-2)(k-4)}{4}(n-1-k\mu)^{k-3} + O(n^{k-4})\Big) \qquad \begin{pmatrix} k \geq 4, \text{ fest} \\ n \to \infty \end{pmatrix}. \end{aligned} \tag{1.4}$$

Aus der Eulerschen Summenformel

$$\sum_{\mu=0}^{m} f(\mu) = \int_0^m f(x)\,dx + \frac{f(0)+f(m)}{2} + \int_0^m \left(x - [x] - \frac{1}{2}\right) f'(x)\,dx \tag{1.5}$$

ergibt sich, wenn $f(x)$ ein Polynom vom Grade $g \geq 1$ ist,

$$\sum_{\mu=0}^{m} f(\mu) = \int_0^m f(x)\,dx + \frac{f(m)}{2} + O(m^{g-1}) \qquad (m \to \infty). \tag{1.6}$$

Auf die erste bzw. zweite Summe von (1.4) wenden wir (1.6) mit

$$f(x) = \left(n - 1 - k\left[\frac{n}{k}\right] + k + kx\right)^g, \quad g = k-2 \quad \text{bzw.} \quad g = k-3$$

an und erhalten

$$\begin{aligned} p_k(n) &= \frac{1}{(k-1)!(k-2)!}\left(\frac{(n-1)^{k-1}}{k(k-1)} + \frac{(n-1)^{k-2}}{2} + \right. \\ &\qquad\qquad \left. + \frac{(k-1)(k-2)(k-4)(n-1)^{k-2}}{4k(k-2)} + O(n^{k-3})\right) \\ &= \frac{1}{k!\,(k-1)!}\left(n^{k-1} + \frac{k(k-1)(k-3)}{4}\, n^{k-2} + O(n^{k-3})\right), \end{aligned}$$

womit Satz 1 bewiesen ist.

Satz 1 läßt sich auch so formulieren:

$$p_k(n) = \frac{1}{k!\,(k-1)!}\left(n + \frac{k(k-3)}{4}\right)^{k-1}(1 + O(n^{-2}))\,, \tag{1.7}$$

$$p_k(n) = \frac{1}{k!}\binom{n-1+\frac{1}{2}\binom{k}{2}}{k-1}(1 + O\,(n^{-2}))\,. \tag{1.8}$$

2. Wir wollen uns jetzt überlegen, wie die weitere Entwicklung von $p_k(n)$ aussieht. Wie man schon an den Fällen $k = 2$ und $k = 3$ sieht, können gewisse Koeffizienten der Entwicklung von der Restklasse von n nach einem gewissen Modul abhängen. So folgt aus (1.1)

$$p_2(n) - \frac{n}{2} = 0, -\frac{1}{2} \qquad \text{für } n \equiv 0, 1 \bmod 2 \tag{2.1}$$

und aus (1.2)

$$p_3(n) - \frac{n^2}{12} = 0, -\frac{1}{12}, -\frac{1}{3}, \frac{1}{4}, -\frac{1}{3}, -\frac{1}{12} \qquad \text{für } n \equiv 0, 1, 2, 3, 4, 5 \bmod 6\,. \tag{2.2}$$

Daß so etwas nicht nur beim konstanten Glied auftritt, zeigt der Fall $k = 4$; aus der bekannten Formel[4])

$$p_4(n) = \frac{1}{36}\left[\frac{n}{2}\right]^2\left(3\left[\frac{n+1}{2}\right] - \left[\frac{n}{2}\right] + 3\right)$$

ergibt sich

$$p_4(n) - \frac{n^3 + 3n^2}{4!\,3!} = 0, \frac{-9n+5}{4!\,3!} \qquad \text{für } n \equiv 0, 1 \bmod 2\,. \tag{2.3}$$

Satz 2. *Es ist*

$$p_k(n) = \frac{1}{(k-1)!\,(k-2)!}\sum_{\varkappa=1}^{k} c_{k\varkappa}(n)\cdot n^{k-\varkappa}\,, \tag{2.4}$$

wo die Koeffizienten $c_{k\varkappa}(n)$ außer von k und $\varkappa$ höchstens noch von der Restklasse von n mod $k!$ abhängen.

Mit anderen Worten: Durchläuft n eine Restklasse mod $k!$, so ist $p_k(n)$ ein Polynom vom Grade $k-1$.

Beweis (durch Induktion nach k). Wegen (1.1) ist Satz 2 richtig für $k = 2$, und wir nehmen an, er sei bereits für $k-1$ bewiesen:

$$p_{k-1}(n) = \frac{1}{(k-2)!\,(k-3)!}\sum_{\nu=1}^{k-1} c_{k-1,\nu}(n)\cdot n^{k-1-\nu} \qquad (k \geqq 3) \tag{2.5}$$

mit

$$c_{k-1,\nu}(n) = c_{k-1,\nu}(m) \qquad \text{für } n \equiv m \bmod (k-1)!\,. \tag{2.6}$$

Nach Hilfssatz 1 und (2.5) ist dann

$$p_k(n) = \frac{1}{(k-1)!\,(k-2)!}\sum_{0 \leqq k\mu \leqq n-k}\;\sum_{\nu=1}^{k-1} c_{k-1,\nu}(n-1-k\mu)\cdot(n-1-k\mu)^{k-1-\nu}.$$

---

[4]) Vgl. [1], 51.

In der Summe über μ summieren wir jetzt getrennt über die einzelnen Restklassen mod $(k-1)!$:

$$\text{(2.7)} \qquad p_k(n) = \frac{1}{(k-1)!\,(k-2)!} \times$$

$$\times \sum_{\nu=1}^{k-1} \sum_{\varrho=0}^{(k-1)!-1} \sum_{0 \le k!\sigma \le n-k-k\varrho} c_{k-1,\nu}(n-1-k\varrho-k!\sigma)\cdot(n-1-k\varrho-k!\sigma)^{k-1-\nu}.$$

Wegen $n-1-k\varrho-k!\sigma \equiv n-1-k\varrho \bmod (k-1)!$ und wegen (2.6) entsteht aus (2.7)

$$\text{(2.8)} \qquad p_k(n) = \frac{1}{(k-1)!\,(k-2)!} \times$$

$$\times \sum_{\nu=1}^{k-1} \sum_{\varrho=0}^{(k-1)!-1} c_{k-1,\nu}(n-1-k\varrho) \sum_{0 \le k!\sigma \le n-k-k\varrho} (n-1-k\varrho-k!\sigma)^{k-1-\nu}.$$

Wir betrachten jetzt die Summe über σ bei festem ν und ϱ; unter der Substitution

$$\text{(2.9)} \qquad n-1-k\varrho = m$$

geht diese über in

$$\sum_{0 \le k!\sigma \le m-k+1} (m-k!\sigma)^{k-1-\nu}.$$

Offenbar ist

$$\sum_{0 \le k!\sigma \le m+k!-k+1} (m+k!-k!\sigma)^{k-1-\nu} - \sum_{0 \le k!\sigma \le m-k+1} (m-k!\sigma)^{k-1-\nu} = (m+k!)^{k-1-\nu}.$$

Daraus folgt: durchläuft m eine feste Restklasse mod $k!$, so durchläuft die erste Differenzenfolge der Summe über σ eine arithmetische Folge der Ordnung $k-1-\nu$. Die Summe über σ durchläuft also eine arithmetische Folge der Ordnung $k-\nu$; also ist

$$\text{(2.10)} \qquad \sum_{0 \le k!\sigma \le m-k+1} (m-k!\sigma)^{k-1-\nu} = \sum_{\lambda=1}^{k-\nu-1} b_{k-\nu,\lambda}\cdot m^{k-\nu-\lambda} + b_{k-\nu,k-\nu}(m),$$

wo die $b_{k-\nu,\lambda}$ nur von k, $k-\nu$ und λ, dagegen $b_{k-\nu,k-\nu}(m)$ nur von k, $k-\nu$ und von der Restklasse von m mod $k!$ abhängt:

$$\text{(2.11)} \qquad b_{k-\nu,k-\nu}(m) = b_{k-\nu,k-\nu}(m') \qquad \text{für } m \equiv m' \bmod k!\,.$$

Setzt man (2.10) mit (2.9) in (2.8) ein, erhält man

$$p_k(n) = \frac{1}{(k-1)!\,(k-2)!} \sum_{\nu=1}^{k-1} \sum_{\varrho=0}^{(k-1)!-1} c_{k-1,\nu}(n-1-k\varrho) \times$$

$$\times \sum_{\lambda=0}^{k-\nu} b_{k-\nu,\lambda}\cdot(n-1-k\varrho)^{k-\nu-\lambda}$$

$$= \frac{1}{(k-1)!\,(k-2)!} \sum_{\nu=1}^{k-1} \sum_{\varrho=0}^{(k-1)!-1} \sum_{\lambda=0}^{k-\nu} c_{k-1,\nu}(n-1-k\varrho)\cdot b_{k-\nu,\lambda} \times$$

$$\times \sum_{\alpha=0}^{k-\nu-\lambda} (-1)^\alpha (1+k\varrho)^\alpha \binom{k-\nu-\lambda}{\alpha} n^{k-\nu-\lambda-\alpha}.$$

Ersetzen wir α durch $\varkappa$ vermöge $\varkappa = \nu + \lambda + \alpha$, so folgt (2.4), wenn wir

$$
\begin{aligned}
\overline{c_{k\varkappa}}(n) = k(k-1) \sum_{\varrho=0}^{(k-1)!-1} \sum_{\substack{\nu \geqq 1 \\ \nu \leqq k-1,\, \nu+\lambda \leqq \varkappa}} \sum_{\lambda \geqq 0} \times \\
\times\, c_{k-1,\nu}(n-1-k\varrho) \cdot b_{k-\nu,\lambda} \cdot (-1)^{\varkappa-\nu-\lambda} (1+k\varrho)^{\varkappa-\nu-\lambda} \binom{k-\nu-\lambda}{\varkappa-\nu-\lambda}
\end{aligned} \tag{2.12}
$$

setzen. Wegen (2.6) und (2.11) haben die $c_{k\varkappa}(n)$ die behauptete Eigenschaft, womit Satz 2 bewiesen ist.

3. Für ein Polynom $f(x)$ vom Grade g läßt sich die Eulersche Summenformel (1.5) auch so schreiben:

$$
\sum_{\tau=1}^{t} f(\tau) = \sum_{\lambda=0}^{g} \frac{B_\lambda}{\lambda!} \left(f^{(\lambda-1)}(t) - f^{(\lambda-1)}(0)\right), \tag{3.1}
$$

wo

$$
B_0 = 1\,, \quad B_1 = \frac{1}{2}\,, \quad f^{(-1)}(t) - f^{(-1)}(0) = \int_0^t f(x)\,dx
$$

gesetzt wurde und wo B_λ die λ-te Bernoullische Zahl bedeutet ($\lambda \geqq 2$). Mit Hilfe von (3.1) kann man die linke Seite von (2.10) berechnen und damit die $b_{k-\nu,\lambda}$ durch die Bernoullischen Zahlen ausdrücken. Dazu haben wir offenbar in (3.1)

$$
f(x) = \left(n - 1 - k\varrho - k!\left[\frac{\left[\frac{n}{k}\right] - 1 - \varrho}{(k-1)!}\right] + k!\,x\right)^{k-1-\nu}, \quad t = \left[\frac{\left[\frac{n}{k}\right] - 1 - \varrho}{(k-1)!}\right]
$$

zu wählen. Wegen

$$
f^{(\lambda-1)}(t) = \frac{(k!)^\lambda}{(k-\nu)\,k!} \left(\prod_{\delta=0}^{\lambda-1} (k-\nu-\delta)\right) (n-1-k\varrho)^{k-\nu-\lambda}
$$

ergibt sich

$$
b_{k-\nu,\lambda} = \frac{(k!)^{\lambda-1} B_\lambda}{(k-\nu)\,\lambda!} \prod_{\delta=0}^{\lambda-1} (k-\nu-\delta)
$$

$$
\begin{aligned}
b_{k-\nu,k-\nu} = -\sum_{\lambda=0}^{k-1-\nu} \frac{(k!)^{\lambda-1} B_\lambda}{(k-\nu)\,\lambda!} \left(\prod_{\delta=0}^{\lambda-1} (k-\nu-\delta)\right) \times \\
\times \left(n - 1 - k\varrho - k!\left[\frac{\left[\frac{n}{k}\right] - 1 - \varrho}{(k-1)!}\right]\right)^{k-\nu-\lambda}.
\end{aligned}
$$

Setzen wir das in (2.12) ein, so sind damit die $c_{k\varkappa}(n)$ mit Hilfe der Bernoullischen Zahlen durch die $c_{k-1,\nu}(n)$ und durch die Restklasse von n mod $k!$ ausgedrückt. Damit sind über (2.12) hinaus alle $c_{k\varkappa}(n)$ berechnet.

4. Wir sind auch in der Lage, die rechte Seite von (0.3) und (0.4) beträchtlich zu verschärfen:

Satz 3. *Für beliebiges $k \geqq 4$ und beliebiges n gilt*

$$
p_k(n) \leqq \frac{1}{k!\,(k-1)!} \left(n + \frac{k\,(k-3)}{4}\right)^{k-1}.
$$

Für reelles $\alpha > 1$ betrachten wir die Funktion $y = x^\alpha$. Für die im Zusammenhang mit dem Mittelwertsatz der Differentialrechnung durch

$$(x+t)^\alpha - x^\alpha = t\alpha(x+t\theta)^{\alpha-1}$$

im Falle $t > 0$ eindeutig festgelegte reelle Größe $\theta = \theta_\alpha(x, t)$ beweist man leicht:

Hilfssatz 2. *Für* $x \geqq 1$, $t > 0$, $\alpha \geqq 2$ *gilt* $\theta_\alpha(x, t) \geqq \frac{1}{2}$.

Beweis von Satz 3 (durch Induktion nach k). Wegen (2.3) ist Satz 3 richtig für $k = 4$ und beliebiges n. Wir nehmen jetzt an, Satz 3 sei schon bewiesen für $k-1$ und beliebiges n ($k \geqq 5$):

$$p_{k-1}(n) \leqq \frac{1}{(k-1)!\,(k-2)!}\left(n + \frac{(k-1)(k-4)}{4}\right)^{k-2}. \tag{4.1}$$

Es ist zu zeigen, daß dann Satz 3 auch für k und beliebiges n gilt; und das geschieht durch Induktion nach n. Zu diesem Zweck machen wir die zusätzliche Annahme, daß Satz 3 auch schon für $p_k(n-k)$ bewiesen sei:

$$p_k(n-k) \leqq \frac{1}{k!\,(k-1)!}\left(n - k + \frac{k(k-3)}{4}\right)^{k-1}. \tag{4.2}$$

Aus Hilfssatz 1, (4.1) und (4.2) folgt dann

$$\begin{aligned} p_k(n) \leqq \frac{1}{(k-1)!\,(k-2)!}&\left(n + \frac{(k-1)(k-4)}{4}\right)^{k-2} + \\ &+ \frac{1}{k!\,(k-1)!}\left(n - k + \frac{k(k-3)}{4}\right)^{k-1}; \end{aligned} \tag{4.3}$$

die rechte Seite von (4.3) ist aber

$$\leqq \frac{1}{k!\,(k-1)!}\left(n + \frac{k(k-3)}{4}\right)^{k-1},$$

wenn man Hilfssatz 2 mit

$$x = n - k + \frac{k(k-3)}{4}, \quad t = k, \quad \alpha = k-1$$

anwendet. Also unter Voraussetzung der Gültigkeit von Satz 3 für $k-1$ und beliebiges n hat sich ergeben, daß aus der Gültigkeit von Satz 3 für k und $n-k$ die Gültigkeit von Satz 3 für k und n folgt. Das ist bereits der Induktionsschritt bezüglich n; es ist jetzt nur noch zu zeigen, daß Satz 3 für k und die kleinsten positiven Reste mod k als Werte für n gilt. Für diese Zahlen, nämlich $n = 1, \ldots, k$ ist aber $p_k(n) \leqq 1$, also Satz 3 trivialerweise gültig. Damit ist Satz 3 bewiesen.

Aus $p_1(n) = 1$, (2.1), (2.2) und Satz 3 ergibt sich

Satz 3. *Für beliebige natürliche Zahlen k und n gilt*

$$p_k(n) \leqq \frac{1}{k!\,(k-1)!}\left(n + \frac{k(k-3)}{4}\right)^{k-1} + 1.$$

Eine entsprechende Verschärfung der linken Seite von (0.3) erfordert ein wenig mehr Aufwand, ist aber mit den hier verwendeten Hilfsmitteln möglich.

5. Neben der Lösungsanzahl $p_k(n)$ von (0.1) ist auch die Lösungsanzahl $q_k(n)$ der Diophantischen Gleichung

$$m_1 + \cdots + m_k = n \qquad (1 \leqq m_1 < \cdots < m_k) \tag{5.1}$$

von Interesse. Setzen wir in (5.1)

$$m_j = n_j + (j-1),$$

erhalten wir

$$n_1 + \cdots + n_k = n - \binom{k}{2} \qquad (1 \leqq n_1 \leqq \cdots \leqq n_k),$$

also

$$q_k(n) = p_k\left(n - \binom{k}{2}\right),$$

womit alles auf das Frühere zurückgeführt ist.

Literatur

[1] Ostmann, H.: Additive Zahlentheorie I. Berlin-Göttingen-Heidelberg: Springer 1956.

(Eingegangen am 27. April 1959)

Christian, U.
Math. Annalen 138, 363—397 (1959)

Über die Multiplikatorensysteme zur Gruppe der ganzen Modulsubstitutionen n-ten Grades

Von

Ulrich Christian in Princeton N. J.

Die verallgemeinerte obere Halbebene $\mathfrak{Z}$, bestehend aus allen n-reihigen komplexen symmetrischen Matrizen $Z = X + iY$ mit positivem Imaginärteil Y, wird durch die Substitution

$$Z^* = U'ZU + S$$

eineindeutig und analytisch auf sich abgebildet; hierbei bedeutet U eine n-reihige unimodulare Matrix, U' ihre Transponierte und S eine n-reihige symmetrische Matrix mit ganzrationalen Elementen. Die Gesamtheit dieser Abbildungen stellt die Gruppe der ganzen Modulsubstitutionen n-ten Grades Δ dar. Es sei $F(Z)$ eine in $\mathfrak{Z}$ holomorphe nicht identisch verschwindende Funktion, die den Bedingungen

$$F(U'ZU + S) = I(U, S; Z)\, F(Z)$$

für alle Elemente aus Δ genügt. Die Symbole $I(U, S; Z)$ bezeichnen hierbei Einheiten, das heißt in $\mathfrak{Z}$ holomorphe Funktionen ohne Nullstellen. Die Gesamtheit der Funktionen $I(U, S; Z)$ heißt das zur Funktion $F(Z)$ gehörige Multiplikatorensystem. Es ist durch $F(Z)$ eindeutig bestimmt. Zu zwei verschiedenen Funktionen $F(Z)$ gehören im allgemeinen verschiedene Multiplikatorensysteme. Die Aufgabe besteht darin, alle möglichen Multiplikatorensysteme anzugeben und in geeigneter Weise zu klassifizieren. Dieses Problem wird für $n \geqq 3$ vollständig gelöst.

Im folgenden werden Matrizen mit großen lateinischen Buchstaben bezeichnet. Der obere Index (a, b) in $M^{(a,b)}$ drücke aus, daß M eine Matrix von a Zeilen und b Spalten ist. $M^{(a)}$ bedeute dasselbe wie $M^{(a,a)}$. Die Symbole M', $\mathrm{Sp}\, M$, $\mathrm{Det}\, M$ benutzen wir für Transponierte, Spur und Determinante einer Matrix $M^{(a)}$; es werde $S^{(a)}[C^{(a,b)}] = C'SC$ definiert. Eine symmetrische Matrix S nenne man positiv bzw. nicht negativ, wenn die quadratische Form $S[x]$ für alle n-reihigen reellen Spalten $x \neq 0$ positiv bzw. nicht negativ ist. $m_{\varkappa\lambda}$ seien die Elemente und $m_{\varkappa\varkappa} = m_\varkappa$ die Diagonalelemente einer Matrix M. Eine Matrix G heiße ganz, wenn ihre sämtlichen Elemente, und halbganz, wenn die Zahlen $g_\varkappa$ und $2g_{\varkappa\lambda}$ ganzrational sind. Unter einer unimodularen Matrix verstehen wir eine ganze Matrix $U^{(a)}$ der Determinante ± 1.

Kapitel I. Multiplikatoren- und Exponentensysteme

§ 1. Grundbegriffe und Probleme

Es sei $\mathfrak{Z}$ eine k-dimensionale komplexe analytische Mannigfaltigkeit und Δ eine Gruppe eineindeutiger holomorpher Abbildungen von $\mathfrak{Z}$ auf sich. Unter einer *Form zur Gruppe* Δ oder kurz *Form* verstehen wir eine in $\mathfrak{Z}$ holomorphe nicht identisch verschwindende Funktion $F(Z)$, welche den Bedingungen

$$F(\Theta Z) = I(\Theta; Z)\, F(Z) \qquad (\Theta \in \Delta) \tag{1}$$

genügt. Hierbei bezeichnet Z den variablen Punkt in $\mathfrak{Z}$ und $I(\Theta; Z)$ für jedes feste $\Theta \in \Delta$ eine Einheit, das bedeutet, eine in $\mathfrak{Z}$ holomorphe Funktion ohne Nullstellen. Die Klasse der durch $F(Z)$ eindeutig bestimmten Einheiten $I(\Theta;Z)$ heißt *das zu $F(Z)$ gehörende Multiplikatorensystem*. Zu verschiedenen Formen gehören im allgemeinen nicht dieselben Multiplikatorensysteme. Der Vielzahl der Formen entspricht eine Gesamtheit von Multiplikatorensystemen, die wir *die zur Gruppe Δ gehörenden Multiplikatorensysteme* oder einfach *die Multiplikatorensysteme* nennen. *Ein Multiplikatorensystem* ist also das zu einer gewissen Form $F(Z)$ gehörende Multiplikatorensystem.

Zwei Formen $F_1(Z)$ und $F_2(Z)$ erkläre man als *äquivalent*, wenn eine Beziehung

$$F_2(Z) = K(Z)\, F_1(Z) \tag{2}$$

mit einer Einheit $K(Z)$ besteht. Sind $I_1(\Theta; Z)$, $I_2(\Theta; Z)$ die zu $F_1(Z)$ bzw. $F_2(Z)$ gehörenden Multiplikatorensysteme, so ergibt sich aus (1) und (2) der Zusammenhang

$$I_2(\Theta; Z) = K(\Theta Z)\, K^{-1}(Z)\, I_1(\Theta; Z) \qquad (\Theta \in \Delta)\,. \tag{3}$$

Wir abstrahieren von den Formen $F_1(Z)$, $F_2(Z)$ und nennen zwei Multiplikatorensysteme $I_1(\Theta; Z)$, $I_2(\Theta; Z)$ *äquivalent*, wenn die Gleichung (3) mit einer passenden Einheit $K(Z)$ erfüllbar ist. Sowohl die Formen als auch die Multiplikatorensysteme fassen wir in Klassen untereinander äquivalenter zusammen.

Für die Theorie der im Raume $\mathfrak{Z}$ definierten Formen zur Gruppe Δ ist wichtig, sämtliche zu Δ gehörenden Multiplikatorensysteme und ihre Einteilung in Klassen untereinander äquivalenter anzugeben. Dieses gelingt, wenn man aus jeder Äquivalenzklasse einen Repräsentanten, etwa das Multiplikatorensystem $I_1(\Theta; Z)$ kennt; denn dann bekommt man die anderen Multiplikatorensysteme $I_2(\Theta; Z)$ derselben Klasse durch die Formeln (3), wobei $K(Z)$ über die Einheiten in $\mathfrak{Z}$ läuft. Ferner ist von Interesse, welche Formen $F(Z)$ dasselbe Multiplikatorensystem haben. Wieder genügt es, die Antwort für einen bestimmten Repräsentanten aus jeder Äquivalenzklasse zu wissen, weil vermöge (2) die Formen $F_1(Z)$ mit dem Multiplikatorensystem $I_1(\Theta; Z)$ in die Gesamtheit der Formen $F_2(Z)$ mit dem Multiplikatorensystem $I_2(\Theta; Z)$ übergehen. Es handelt sich also darum, aus jeder Klasse untereinander äquivalenter Multiplikatorensysteme in möglichst einfacher Weise genau einen Repräsentanten anzugeben. Dieses Problem ist eng verwandt

mit der Frage nach allen Systemen von Einheiten $E(\Theta; Z)$ $(\Theta \in \Delta)$, die in der Form

$$D(\Theta Z) D^{-1}(Z) = E(\Theta; Z) \qquad (\Theta \in \Delta) \tag{4}$$

durch eine von Θ unabhängige Einheit $D(Z)$ ausdrückbar sind.

Bei einer einfach zusammenhängenden Mannigfaltigkeit $\mathfrak{Z}$ sind die Ausdrücke

$$D(Z) = e^{2\pi i M(Z)},$$

mit einer holomorphen Funktion $M(Z)$, sämtliche Einheiten. Die Funktion $M(Z)$ ist durch Vorgabe von $D(Z)$ bis auf eine ganzrationale additive Konstante eindeutig bestimmt. Wir schreiben die Multiplikatorensysteme in der Gestalt

$$I(\Theta; Z) = e^{2\pi i J(\Theta; Z)} \qquad (\Theta \in \Delta), \tag{5}$$

nennen die Funktionen $J(\Theta; Z)$ *das zur Funktion $F(Z)$ gehörende Exponentensystem* und die Gesamtheit der Exponentensysteme *die zur Gruppe Δ gehörenden Exponentensysteme*. Zu beachten ist, daß bei jedem Exponenten $J(\Theta; Z)$ eine ganzrationale additive Konstante willkürlich bleibt. Setzen wir noch

$$K(Z) = e^{2\pi i L(Z)}, \; E(\Theta; Z) = e^{2\pi i N(\Theta; Z)},$$

so ergeben die Gleichungen (3), (4) die Relationen

$$J_2(\Theta; Z) \equiv L(\Theta Z) - L(Z) + J_1(\Theta; Z) \qquad (\Theta \in \Delta) \tag{6}$$

bzw.

$$M(\Theta Z) - M(Z) \equiv N(\Theta; Z) \qquad (\Theta \in \Delta).$$

Das Kongruenzzeichen bedeutet, daß beide Seiten sich nur um eine ganzrationale additive Konstante unterscheiden. Zwei Exponentensysteme $J_1(\Theta; Z)$ und $J_2(\Theta; Z)$ sind *äquivalent*, wenn die Kongruenz (6) mit einer passenden in $\mathfrak{Z}$ holomorphen Funktion $L(Z)$ gilt. Die Auffindung eines Repräsentanten aus jeder Äquivalenzklasse von Multiplikatorensystemen ist mit der Angabe eines Elements aus jeder Klasse äquivalenter Exponentensysteme identisch. Letzteres führt auf die Untersuchung gewisser Differenzengleichungen.

In unserem Falle ist $\mathfrak{Z}$ die $k = \frac{n(n+1)}{2}$-dimensionale verallgemeinerte obere Halbebene, bestehend aus allen komplexen, symmetrischen Matrizen $Z^{(n)} = X + iY$ mit positivem Imaginärteil Y und Δ die Gruppe der ganzen Modulsubstitutionen n-ten Grades

$$Z^* = Z[U] + S;$$

hierbei bedeuten $U^{(n)}$ eine beliebige unimodulare sowie $S^{(n)}$ eine ganze symmetrische Matrix. Die Multiplikatoren- und Exponentensysteme schreiben wir in der Form $I(U, S; Z)$ bzw. $J(U, S; Z)$, und die Gleichungen (1), (5) nehmen die Gestalt

$$F(Z[U] + S) = I(U, S; Z) F(Z) \tag{7}$$

bzw.

$$I(U, S; Z) = e^{2\pi i J(U, S; Z)} \tag{8}$$

an. Die Aufgabe, aus jeder Klasse äquivalenter Multiplikatorensysteme einen Repräsentanten anzugeben, wird für $n \geqq 3$ gelöst.

§ 2. *Relationen*

Nicht jede Menge von Einheiten $I(\Theta; Z)$ $(\Theta \in \Delta)$ kann ein Multiplikatorensystem sein. Sind nämlich Θ_1, Θ_2 zwei Abbildungen aus Δ, so erhält man gemäß (1) die Gleichungen

$$F(\Theta_2\Theta_1 Z) = I(\Theta_2\Theta_1; Z)\, F(Z) = I(\Theta_2; \Theta_1 Z)\, I(\Theta_1; Z)\, F(Z)$$

und hieraus

$$I(\Theta_2\Theta_1; Z) = I(\Theta_2; \Theta_1 Z)\, I(\Theta_1; Z) \qquad (\Theta_1, \Theta_2 \in \Delta)\,. \tag{9}$$

Diesen Relationen müssen sämtliche Einheiten eines Multiplikatorensystems genügen. In der Schreibweise der Exponentensysteme nehmen die Formeln (9) die Gestalt

$$J(\Theta_2\Theta_1; Z) \equiv J(\Theta_2; \Theta_1 Z) + J(\Theta_1; Z) \qquad (\Theta_1, \Theta_2 \in \Delta) \tag{10}$$

an.

Kapitel II. Differenzengleichungen

§ 1. *Das Problem*

Es sei $\mathfrak{Z}$ eine komplexe analytische Mannigfaltigkeit und Δ eine Gruppe eineindeutiger holomorpher Abbildungen von $\mathfrak{Z}$ auf sich. Jedem Element $\Theta \in \Delta$ werde eine in $\mathfrak{Z}$ holomorphe Funktion $N(\Theta; Z)$ zugeordnet. Wir fragen, für welche Systeme von Funktionen $N(\Theta; Z)$ $(\Theta \in \Delta)$ die Differenzengleichungen

$$M(\Theta Z) - M(Z) = N(\Theta; Z) \qquad (\Theta \in \Delta) \tag{11}$$

eine in $\mathfrak{Z}$ holomorphe Lösung $M(Z)$ besitzen. Bei Bestehen der Gleichung (11) folgt für irgend zwei Elemente $\Theta_1, \Theta_2 \in \Delta$ durch Umrechnung

$$\begin{aligned} N(\Theta_2\Theta_1; Z) = M(\Theta_2\Theta_1 Z) - M(Z) &= M(\Theta_2\Theta_1 Z) - M(\Theta_1 Z) + M(\Theta_1 Z) \\ &- M(Z) = N(\Theta_2; \Theta_1 Z) + N(\Theta_1; Z)\,. \end{aligned}$$

Die Bedingungen

$$N(\Theta_2\Theta_1; Z) = N(\Theta_2; \Theta_1 Z) + N(\Theta_1; Z) \tag{12}$$

sind also für die Lösbarkeit der Differenzengleichungen (11) notwendig. Ob sie genügen, untersuchen wir für den Fall, daß $\mathfrak{Z}$ die verallgemeinerte obere Halbebene und Δ die Gruppe der ganzen Modulsubstitutionen n-ten Grades ist. Das Differenzengleichungssystem lautet dann

$$M(Z[U] + S) - M(Z) = N(U, S; Z)\,, \tag{13}$$

und die Formeln (12) nehmen die Gestalt

$$N(U_1U_2, S_1[U_2] + S_2; Z) = N(U_2, S_2; Z[U_1] + S_1) + N(U_1, S_1; Z) \tag{14}$$

an. Hauptziel dieses Kapitels ist der Beweis von

Satz 1: *Die Relationen* (14) *sind für die Lösbarkeit der Differenzengleichungen* (13) *notwendig und hinreichend* ($n \neq 2$).

Sowohl zum Beweise von Satz 1 als auch für die später folgende Untersuchung der Exponentensysteme erweist es sich als nützlich, zuerst das Differenzengleichungssystem

$$M(Z+S)-M(Z)=N(E,S;Z)$$

zu betrachten. E bedeutet dabei die Einheitsmatrix. Hierfür ordnen wir die Variablen zu einer Spalte z an, deren Zeilenzahl eine beliebige natürliche Zahl m sein kann, bezeichnen mit $\mathfrak{Z}=\mathfrak{Z}(\mathfrak{Y})$ die Gesamtheit aller Punkte $z=x+iy$, für welche der Imaginärteil y in einem vorgegebenen Bereich $\mathfrak{Y}$ des m-dimensionalen reellen y-Raumes liegt und schreiben die zu untersuchenden Differenzengleichungen in der Form

$$G(z+s)-G(z)=H(s,z)\,; \tag{15}$$

s bedeutet hierbei eine ganze m-zeilige Spalte. Die Beziehungen (12) lauten jetzt

$$H(s_1+s_2;z)=H(s_2;z+s_1)+H(s_1;z)\,. \tag{16}$$

Es gilt

Satz 2: *Ist* $\mathfrak{Y}$ *eine offene konvexe Pyramide, so sind die Relationen* (16) *für die Lösbarkeit der Differenzengleichungen* (15) *durch eine in* $\mathfrak{Z}(\mathfrak{Y})$ *holomorphe Funktion* $G(z)$ *notwendig und hinreichend.*

Die Gesamtheit der reellen symmetrischen Matrizen $Y>0$ bildet eine konvexe Pyramide, und Satz 2 läßt sich daher auf die verallgemeinerte obere Halbebene $\mathfrak{Z}$ anwenden. Für $n=1$ besteht die unimodulare Gruppe nur aus den Elementen ± 1, so daß Satz 1 aus Satz 2 folgt. Im weiteren kann somit $n\geqq 2$ vorausgesetzt werden. Wegen der Gleichungen (14) erfüllen die Funktionen $N(E,S;Z)$ die Bedingungen (16), sind gemäß Satz 2 in der Gestalt

$$G(Z+S)-G(Z)=N(E,S;Z) \tag{17}$$

darstellbar. Diese Gleichungen (17) bleiben richtig, wenn man zu $G(Z)$ eine beliebige in $\mathfrak{Z}$ holomorphe Funktion mit den Perioden S addiert. Um die Differenzengleichungen (13) zu lösen, machen wir daher den Ansatz

$$M(Z)=G(Z)+P(Z)$$

mit einer in $\mathfrak{Z}$ holomorphen Funktion $P(Z)$, welche die Periodizitätseigenschaft

$$P(Z+S)-P(Z)=0 \tag{18}$$

besitzt, und haben nunmehr die Gleichungen

$$P(Z[U])-P(Z)=Q(U;Z) \tag{19}$$

mit

$$Q(U;Z)=N(U,S;Z)-G(Z[U]+S)+G(Z)$$

durch eine geeignete Funktion $P(Z)$ zu befriedigen. Hierzu muß man als erstes zeigen, daß die Funktionen $Q(U;Z)$ wohldefiniert, d. h. von S unabhängig, und ferner die von (18), (12) herrührenden notwendigen Relationen

$$Q(U;Z+S)=Q(U;Z) \tag{20}$$

bzw.

$$Q(U_1U_2;Z)=Q(U_2;Z[U_1])+Q(U_1;Z) \tag{21}$$

richtig sind. Vermöge der Beziehungen (14), (17) ergibt sich aber tatsächlich

$$Q(U;Z) = N(U,S;Z) - G(Z[U]+S) + G(Z) = N(U,S;Z) -$$
$$- N(E,S;Z[U]) - G(Z[U]) + G(Z) = N(U,O;Z) - G(Z[U]) + G(Z),$$

$$\begin{aligned} Q(U;Z+S) &= N(U,O;Z+S) - G(Z[U]+S[U]) + G(Z+S) \\ &= N(U,O;Z+S) + N(E,S;Z) - N(E,S[U];Z[U]) - G(Z[U]) + G(Z) \\ &= N(U,S[U];Z) - N(E,S[U];Z[U]) - G(Z[U]) + G(Z) \\ &= N(U,O;Z) - G(Z[U]) + G(Z) = Q(U;Z), \end{aligned}$$

$$\begin{aligned} Q(U_1U_2;Z) &= N(U_1U_2,O;Z) - G(Z[U_1U_2]) + G(Z) \\ &= N(U_2,O;Z[U_1]) - G(Z[U_1U_2]) + G(Z[U_1]) + N(U_1,O;Z) - G(Z[U_1]) + \\ &\quad + G(Z) = Q(U_2;Z[U_1]) + Q(U_1;Z). \end{aligned}$$

Satz 1 ist damit zurückgeführt auf den Beweis von

Satz 3: *Für die Lösbarkeit der Differenzengleichungen* (18), (19) *sind die Relationen* (20), (21) *notwendig und hinreichend* ($n \geqq 3$).

§ 2. Beweis von Satz 2

Die linken Seiten der Gleichungen (16) ändern sich bei Vertauschung von s_1 und s_2 nicht. Daraus folgt

$$H(s_1;z) + H(s_2;z+s_1) = H(s_2;z) + H(s_1;z+s_2),$$

also insbesondere

$$H_\mu(z) + H_\nu(z+e_\mu) = H_\nu(z) + H_\mu(z+e_\nu) \qquad (\mu,\nu = 1,\ldots,m), \tag{22}$$

wobei e_μ die μ-te Einheitsspalte von m Zeilen bezeichnet und $H_\mu(z) = H(e_\mu;z)$ gesetzt wurde. Unter alleiniger Benutzung der Relationen (22) werden wir die m Differenzengleichungen

$$G(z+e_\mu) - G(z) = H_\mu(z) \qquad (\mu = 1,\ldots,m) \tag{23}$$

durch eine in $\mathfrak{Z}(\mathfrak{Y})$ holomorphe Funktion $G(z)$ lösen. Diese erfüllt die Gleichungen (15) auf Grund des Zusammenhangs (16).

Hilfssatz 1: *Jede in der komplexen w-Ebene holomorphe Funktion* $E(w)$ *gestattet eine Darstellung*

$$D(w+1) - D(w) = E(w)$$

mit einer in der w-Ebene holomorphen Funktion $D(w)$.

Dieser Hilfssatz ist bekannt [1, 4, 5, 6] und wird hier nicht bewiesen.

Hilfssatz 2: *Es sei* $\mathfrak{Y}$ *eine offene Punktmenge des reellen m-dimensionalen y-Raums; in* $\mathfrak{Z}(\mathfrak{Y})$ *seien m holomorphe Funktionen* $H_1(z),\ldots,H_m(z)$ *erklärt.* $\mathfrak{Y}$ *lasse sich durch eine monoton wachsende Folge offener Bereiche* $\mathfrak{Y}_\tau$ ($\tau = 1,2,\ldots$), *welche die Eigenschaft besitzen, daß in* $\mathfrak{Z}(\mathfrak{Y}_\tau)$ *eine holomorphe Funktion* $G_\tau(z)$ *mit*

$$G_\tau(z+e_\mu) - G_\tau(z) = H_\mu(z) \qquad (\mu = 1,\ldots,m) \tag{24}$$

existiert, ausschöpfen. Dann sind die Differenzengleichungen (23) *durch eine in* $\mathfrak{Z}(\mathfrak{Y})$ *holomorphe Funktion* $G(z)$ *lösbar.*

Beweis: Man wähle eine Folge kompakter Bereiche $\mathfrak{R}_\tau$ des y-Raums, so daß

$$\mathfrak{R}_\tau \subset \mathfrak{Y}_\tau,\ \mathfrak{R}_\tau \subset \mathfrak{R}_{\tau+1}(\tau = 1, 2, \ldots);\ \bigcup_\tau \mathfrak{R}_\tau = \mathfrak{Y}$$

gilt. Die in $\mathfrak{Z}(\mathfrak{Y}_\tau)$ erklärte Funktion

$$P_\tau(z) = G_{\tau+1}(z) - G_\tau(z)$$

hat auf Grund von (24) die Perioden $e_1, \ldots, e_m$ und läßt sich daher in eine m-fache Fourierreihe entwickeln. Wegen der Kompaktheit von $\mathfrak{R}_\tau$ konvergiert diese Fourierreihe in $\mathfrak{Z}(\mathfrak{R}_\tau)$ gleichmäßig. Folglich ist $P_\tau(z)$ die Summe einer für alle komplexen Vektoren z holomorphen endlichen Partialsumme $Q_\tau(z)$ und eines Restes $R_\tau(z)$, welcher der Abschätzung

$$|R_\tau(z)| \leqq \frac{1}{\tau^2} \qquad (z \in \mathfrak{Z}(\mathfrak{R}_\tau))$$

genügt. Der Grenzwert

$$G(z) = \lim_{\varrho\to\infty}\left(G_\varrho(z) - \sum_{\sigma=1}^{\varrho-1} Q_\sigma(z)\right)$$

existiert auf jedem der Bereiche $\mathfrak{Z}(\mathfrak{R}_\tau)$ $(\tau = 1, 2, \ldots)$ gleichmäßig und stellt in

$$\bigcup_\tau \mathfrak{Z}(\mathfrak{R}_\tau) = \mathfrak{Z}(\mathfrak{Y})$$

eine holomorphe Funktion dar. Die Partialsummen $Q_\sigma(z)$ haben die Perioden $e_1, \ldots, e_m$; aus den Formeln (24) ergeben sich somit die Gleichungen (23) für die hier konstruierte Funktion $G(z)$.

Von zwei l-zeiligen reellen Vektoren p, q wollen wir sagen, daß $p > q$ ist, wenn $p - q$ positive Komponenten hat.

Hilfssatz 3: *Falls $\mathfrak{Y}$ das Intervall $a < y < b$ darstellt und die Relationen* (22) *gelten, ist das Differenzengleichungssystem* (23) *in $\mathfrak{Z}(\mathfrak{Y})$ lösbar. Dabei bedeuten a, b Vektoren, deren m Komponenten reell oder unendlich sind.*

Beweis: Es sei

$$a = \begin{pmatrix} a_1 \\ \vdots \\ a_m \end{pmatrix},\quad b = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix},\quad c = \begin{pmatrix} c_1 \\ \vdots \\ c_m \end{pmatrix},\quad d = \begin{pmatrix} d_1 \\ \vdots \\ d_m \end{pmatrix},\quad y = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix},\quad z = \begin{pmatrix} z_1 \\ \vdots \\ z_m \end{pmatrix},\ .$$

$$a < c < d < b$$

und $\mathfrak{T}_\mu (1 \leqq \mu \leqq m)$ die Menge aller Vektoren y, die den Ungleichungen

$$c_\nu < y_\nu < d_\nu (\nu = 1, \ldots, \mu);\quad a_\nu < y_\nu < b_\nu (\nu = \mu + 1, \ldots, m)$$

genügen. Wir wollen in $\mathfrak{Z}(\mathfrak{T}_1)$ eine holomorphe Funktion $G_1(z)$, welche die Differenzengleichung

$$G_1(z + e_1) - G_1(z) = H_1(z)$$

löst, konstruieren. Unter Benutzung des Mittag-Lefflerschen Satzes bestimmen wir in der komplexen w-Ebene eine meromorphe Funktion $C(w)$ mit den Hauptteilen

$$(25) \qquad \frac{1}{w - \gamma} \qquad (\gamma = 1, 2, \ldots).$$

Der Ausdruck

$$E(w) = \frac{1}{w} + C(w) - C(w+1)$$

ist ganz, gestattet also nach Hilfssatz 1 eine Darstellung

$$E(w) = D(w+1) - D(w)$$

mit ganzem $D(w)$. Folglich hat

$$A(w) = C(w) + D(w)$$

die in (25) angegebenen Hauptteile; ferner gilt

$$A(w+1) - A(w) = \frac{1}{w}. \tag{26}$$

Auf dieselbe Art erhalten wir eine meromorphe Funktion $B(w)$, deren Hauptteile durch

$$\frac{-1}{w+\gamma} \qquad (\gamma = 0, 1, 2, \ldots) \tag{27}$$

gegeben sind und für die

$$B(w+1) - B(w) = \frac{1}{w} \tag{28}$$

ist. Bei reellem $t > 1$ bedeute $\mathfrak{A}$ den Polygonzug

$$-t + id_1, t + id_1, t + ic_1, -t + ic_1;$$

$\mathfrak{B}$ die Strecke von $-t + ic_1$ nach $-t + id_1$; ferner $\mathfrak{G}(t)$ das Rechteck

$$z_1 = x_1 + iy_1, -t < x_1 < t - 1, c_1 < y_1 < d_1$$

und $\mathfrak{H}(t)$ das Rechteck

$$z_1 = x_1 + iy_1, -t < x_1 < t, c_1 < y_1 < d_1$$

in der z_1-Ebene. Dann liefert der Cauchysche Integralsatz

$$H_1(z) = \frac{1}{2\pi i} \int\limits_{\mathfrak{A}\mathfrak{B}} \frac{H_1(u, z_2, \ldots, z_m)}{z_1 - u}\, du \qquad (z_1 \in \mathfrak{H}(t), z \in \mathfrak{Z}(\mathfrak{T}_1)). \tag{29}$$

Die Summe

$$F(t; z) = \frac{1}{2\pi i}\left(\int\limits_{\mathfrak{A}} A(z_1 - u)\, H_1(u, z_2, \ldots, z_m)\, du + \int\limits_{\mathfrak{B}} B(z_1 - u)\, H_1(u, z_2, \ldots, z_m)\, du\right) \tag{30}$$

erweist sich für $z_1 \in \mathfrak{H}(t)$, $z \in \mathfrak{Z}(\mathfrak{T}_1)$ unter Benutzung von (25), (27) als holomorph, und aus (26), (28) bis (30) folgt

$$F(t; z + e_1) - F(t; z) = H_1(z) \qquad (z_1 \in \mathfrak{G}(t), z \in \mathfrak{Z}(\mathfrak{T}_1)). \tag{31}$$

Weil die Differenz

$$F(2; z) - F(t; z) = P(z) \qquad (t > 1, z_1 \in \mathfrak{H}(t) \cap \mathfrak{H}(2), z \in \mathfrak{Z}(\mathfrak{T}_1))$$

die Periode e_1 hat, läßt sie sich auf den Bereich $\mathfrak{Z}(\mathfrak{T}_1))$ holomorph fortsetzen; daraus ergibt sich die holomorphe Fortsetzbarkeit der Funktion

$$G_1(z) = F(2; z)$$

auf $\mathfrak{Z}(\mathfrak{T}_1)$, und gemäß (31) ist

$$G_1(z+e_1) - G_1(z) = H_1(z) \qquad (z \in \mathfrak{Z}(\mathfrak{T}_1)).$$

Aus (30) lesen wir noch eine Eigenschaft von $G_1(z)$ ab, die wir später benötigen. Hat nämlich $H_1(z)$ in einigen der Variablen $z_2, \ldots, z_m$ die Periode 1, so auch die Funktion $G_1(z)$.

Unter der Annahme, daß wir in $\mathfrak{Z}(\mathfrak{T}_\mu)$ bereits eine holomorphe Funktion $G_\mu(z)$ mit

$$(32) \qquad G_\mu(z+e_\nu) - G_\mu(z) = H_\nu(z) \qquad (\nu = 1, \ldots, \mu)$$

gefunden haben, beweisen wir für den Fall $\mu < m$ die Existenz einer in $\mathfrak{Z}(\mathfrak{T}_{\mu+1})$ holomorphen Funktion $G_{\mu+1}(z)$, welche die Bedingungen

$$(33) \qquad G_{\mu+1}(z+e_\nu) - G_{\mu+1}(z) = H_\nu(z) \qquad (\nu = 1, \ldots, \mu+1)$$

befriedigt. Wir bilden

$$(34) \qquad K(z) = H_{\mu+1}(z) - G_\mu(z+e_{\mu+1}) + G_\mu(z) \qquad (z \in \mathfrak{Z}(\mathfrak{T}_\mu))$$

und leiten aus (22), (32) die Periodenrelationen

$$K(z+e_\nu) - K(z) = 0 \qquad (\nu = 1, \ldots, \mu)$$

her. Nun lösen wir nach der früheren Methode in $\mathfrak{Z}(\mathfrak{T}_{\mu+1})$ die Differenzengleichung

$$(35) \qquad F(z+e_{\mu+1}) - F(z) = K(z);$$

gemäß unserer Anmerkung ergibt sich zusätzlich

$$(36) \qquad F(z+e_\nu) - F(z) = 0 \qquad (\nu = 1, \ldots, \mu).$$

Wegen (32), (34) bis (36) erfüllt

$$G_{\mu+1}(z) = G_\mu(z) + F(z)$$

die Beziehungen (33). Induktiv haben wir also die Differenzengleichungen (23) in $\mathfrak{Z}(\mathfrak{T})$ gelöst, wobei $\mathfrak{T} = \mathfrak{T}_m$ das Intervall $c < y < d$ bezeichnet.

Strebt c fallend gegen a und d wachsend gegen b, so durchläuft $\mathfrak{T}$ eine monotone Folge offener Bereiche, die $\mathfrak{Y}$ ausschöpfen. Auf Grund von Hilfssatz 2 ist Hilfssatz 3 bewiesen.

Hilfssatz 4: *Es seien a, b wie in Hilfssatz 3 erklärt, R eine m-reihige reelle Matrix mit nicht verschwindender Determinante, $\mathfrak{Y}$ der Bereich $a < R^{-1}y < b$ und die Bedingungen (22) erfüllt. Die Differenzengleichungen (23) sind dann in $\mathfrak{Z}(\mathfrak{Y})$ lösbar.*

Beweis: Die Richtigkeit des Hilfssatzes 4 wird zuerst für alle Matrizen

$$(37) \qquad R = (r_1, \ldots, r_\tau, e_{\tau+1}, \ldots, e_m) \qquad (0 \leq \tau \leq m)$$

mit rationalen Spalten $r_1, \ldots, r_\tau$ durch Induktion nach τ gezeigt. Bei $\tau = 0$ ist R die Einheitsmatrix, es liegt also der in Hilfssatz 3 behandelte Sachverhalt vor. Jetzt sei $\tau > 0$ und die Behauptung für $\tau - 1$ statt τ bereits bewiesen.

Um die Gültigkeit von Hilfssatz 4 für τ darzulegen, beachten wir zunächst, daß wegen $(\operatorname{Det} R) \neq 0$ die Determinante des τ-reihigen linken oberen Kästchens von R ungleich Null ist. Unter geeigneter Numerierung der Vektoren $r_1, \ldots, r_\tau$ sowie der Zeilen von a, b verschwindet folglich die τ-te Komponente der Spalte r_τ nicht. Wir führen den positiven Hauptnenner g von r_τ, den Vektor

$$s = g r_\tau = \begin{pmatrix} s_1 \\ \vdots \\ s_m \end{pmatrix},$$

die Matrizen

$$S = (r_1, \ldots, r_{\tau-1}, s, e_{\tau+1}, \ldots, e_m),$$
$$U = R^{-1} S = (e_1, \ldots, e_{\tau-1}, g e_\tau, e_{\tau+1}, \ldots, e_m),$$
$$T = (e_1, \ldots, e_{\tau-1}, s, e_{\tau+1}, \ldots, e_m), \tag{38}$$

die Variablen

$$w = T^{-1} z \tag{39}$$

sowie die Funktionen

$$K_\mu(w) = H_\mu(Tw) \qquad (\mu \neq \tau;\ \mu = 1, \ldots, m), \tag{40}$$

$$K_\tau(w) = \sum_{\substack{\lambda=1\\ \lambda \neq \tau}}^{m} \left[(\operatorname{sign} s_\lambda) \sum_{\sigma=0}^{|s_\lambda|-1} H_\lambda\left(Tw + \sum_{\substack{\varrho=1\\ \varrho\neq\tau}}^{\lambda-1} s_\varrho e_\varrho + s_\tau e_\tau + \frac{s_\lambda - |s_\lambda|}{2} e_\lambda + \sigma e_\lambda \right)\right] + (\operatorname{sign} s_\tau) \sum_{\sigma=0}^{|s_\tau|-1} H_\tau\left(Tw + \frac{s_\tau - |s_\tau|}{2} e_\tau + \sigma e_\tau \right) \tag{41}$$

ein; durch elementare Rechnung leiten sich aus (22), (39) die Relationen

$$K_\mu(w) + K_\nu(w + e_\mu) = K_\nu(w) + K_\mu(w + e_\nu) \qquad (\mu, \nu = 1, \ldots, m) \tag{42}$$

her. Der Imaginärteil

$$v = T^{-1} y$$

von w durchläuft den Bereich

$$U^{-1} a < (T^{-1} S)^{-1} v < U^{-1} b,$$

falls y in

$$a < R^{-1} y < b$$

variiert. Weil

$$T^{-1} S = (f_1, \ldots, f_{\tau-1}, e_\tau, e_{\tau+1}, \ldots, e_m)$$

rationale Spalten $f_1, \ldots, f_{\tau-1}$ hat, wird die Induktionsvoraussetzung anwendbar; es folgt auf Grund von (42) die Existenz einer holomorphen Funktion $L(w)$, welche die Bedingungen

$$L(w + e_\mu) - L(w) = K_\mu(w) \qquad (\mu = 1, \ldots, m)$$

erfüllt. Für

$$F(z) = L(T^{-1} z) \qquad (z \in \mathfrak{Z}(\mathfrak{Y}))$$

gehen diese Gleichungen gemäß (38) bis (41) in

$$F(z + e_\mu) - F(z) = H_\mu(z) \qquad (\mu \neq \tau;\ \mu = 1, \ldots, m), \tag{43}$$

$$F(z+s)-F(z)$$
$$=\sum_{\substack{\lambda=1\\ \lambda\neq\tau}}^{m}\left[(\operatorname{sign}s_\lambda)\sum_{\sigma=0}^{|s_\lambda|-1}H_\lambda\left(z+\sum_{\substack{\varrho=1\\ \varrho\neq\tau}}^{\lambda-1}s_\varrho e_\varrho+s_\tau e_\tau+\frac{s_\lambda-|s_\lambda|}{2}e_\lambda+\sigma e_\lambda\right)\right]+$$
$$+(\operatorname{sign}s_\tau)\sum_{\sigma=0}^{|s_\tau|-1}H_\tau\left(z+\frac{s_\tau-|s_\tau|}{2}e_\tau+\sigma e_\tau\right)$$

über, woraus sich unter Benutzung der trivialen Identität

$$F(z+s)-F(z)$$
$$=\sum_{\substack{\lambda=1\\ \lambda\neq\tau}}^{m}\left[(\operatorname{sign}s_\lambda)\sum_{\sigma=0}^{|s_\lambda|-1}\left\{F\left(z+\sum_{\substack{\varrho=1\\ \varrho\neq\tau}}^{\lambda-1}s_\varrho e_\varrho+s_\tau e_\tau+\frac{s_\lambda-|s_\lambda|}{2}e_\lambda+\sigma e_\lambda+e_\lambda\right)-\right.\right.$$
$$\left.\left.-F\left(z+\sum_{\substack{\varrho=1\\ \varrho\neq\tau}}^{\lambda-1}s_\varrho e_\varrho+s_\tau e_\tau+\frac{s_\lambda-|s_\lambda|}{2}e_\lambda+\sigma e_\lambda\right)\right\}\right]+F(z+s_\tau e_\tau)-F(z)$$

die Beziehung

$$F(z+he_\tau)-F(z)=\sum_{\sigma=0}^{h-1}H_\tau(z+\sigma e_\tau)\qquad(z\in\mathfrak{Z}(\mathfrak{Y}))\tag{44}$$

ergibt; hierbei wurde $h=|s_\tau|$ gesetzt. h ist positiv, da die τ-te Komponente des Vektors r_τ nicht verschwindet.

Zufolge der Formeln (22), (43) besitzt

$$M(z)=H_\tau(z)-F(z+e_\tau)+F(z)\qquad(z\in\mathfrak{Z}(\mathfrak{Y}))\tag{45}$$

die Perioden $e_\mu(\mu\neq\tau;\mu=1,\ldots,m)$. Aus (44) bekommen wir

$$M(z+he_\tau)=M(z)\qquad(z\in\mathfrak{Z}(\mathfrak{Y})),$$

also die Möglichkeit einer Fourierschen Entwicklung

$$M(z)=\sum_{\nu=-\infty}^{\infty}P_\nu(z_1,\ldots,z_{\tau-1},z_{\tau+1},\ldots,z_m)\,e^{2\pi i\frac{\nu}{h}z_\tau}=\sum_{\eta=0}^{h-1}e^{2\pi i\frac{\eta}{h}z_\tau}M_\eta(z).$$

Die auf $\mathfrak{Z}(\mathfrak{Y})$ holomorphen Summen

$$M_\eta(z)=\sum_{\nu=-\infty}^{\infty}P_{h\nu+\eta}(z_1,\ldots,z_{\tau-1},z_{\tau+1},\ldots,z_m)\,e^{2\pi i\nu z_\tau}\qquad(\eta=0,\ldots,h-1)$$

haben die Perioden $e_1,\ldots,e_m$. Wählt man nun gemäß Hilfssatz 1 in der komplexen u-Ebene ganze Funktionen $D_0(u),\ldots,D_{h-1}(u)$ mit

$$D_\eta(u+1)-D_\eta(u)=e^{2\pi i\frac{\eta}{h}u}\qquad(\eta=0,\ldots,h-1),$$

so genügen die Ausdrücke

$$N(z)=\sum_{\eta=0}^{h-1}D_\eta(z_\tau)\,M_\eta(z)\qquad(z\in\mathfrak{Z}(\mathfrak{Y}))$$

den Bedingungen

$$N(z+e_\tau)-N(z)=M(z),\ N(z+e_\mu)-N(z)=0\qquad(\mu\neq\tau;\mu=1,\ldots,m),$$

und

$$G(z)=F(z)+N(z)$$

befriedigt das Differenzengleichungssystem (23). Damit ist Hilfssatz 4 für rationale Matrizen R bewiesen.

Die reellen Matrizen R können wir jetzt ganz leicht erfassen. Es seien a_λ, b_λ ($\lambda = 1, 2, \ldots$) gegen a bzw. b strebende Folgen von Vektoren, R_λ nicht ausgeartete rationale Matrizen mit

$$\lim_{\lambda \to \infty} R_\lambda = R\,,$$

$\mathfrak{Y}_\lambda$ das Gebiet

$$a_\lambda < R_\lambda^{-1} y < b_\lambda \qquad (\lambda = 1, 2, \ldots).$$

Bei passender Wahl der Folgen a_λ, b_λ, R_λ erreichen wir

$$\mathfrak{Y}_1 \subset \mathfrak{Y}_2 \subset \cdots \subset \mathfrak{Y}\,, \quad \bigcup_\nu \mathfrak{Y}_\lambda = \mathfrak{Y}\,,$$

und Hilfssatz 4 folgt wegen Hilfssatz 2 aus der Lösbarkeit der Differenzengleichungen (23) in jedem der Bereiche $\mathfrak{Z}(\mathfrak{Y}_\lambda)$ ($\lambda = 1, 2, \ldots$).

Hilfssatz 5: *Die Differenzengleichungen (23) lassen sich unter Voraussetzung von (22) in $\mathfrak{Z}(\mathfrak{Y})$ lösen, wenn $\mathfrak{Y}$ durch*

$$y = Rt,\ a < t < b$$

definiert ist. Dabei bedeuten a, b, t Vektoren von l Zeilen ($l \geq m$), deren Komponenten reell oder unendlich sind. $R^{(m,l)}$ stellt eine reelle Matrix vom Range m dar.

Beweis: Wir wenden Induktion nach $h = l - m$ an. Für $h = 0$ liegt der in Hilfssatz 4 genannte Sachverhalt vor. Ist $h > 0$, wählen wir eine l-spaltige reelle Zeile d, so daß die Matrix

$$\begin{pmatrix} R \\ d \end{pmatrix}$$

den Rang $m + 1$ hat, und eine komplexe Variable $w = u + iv$; ferner benötigen wir den durch

$$\begin{pmatrix} y \\ v \end{pmatrix} = \begin{pmatrix} R \\ d \end{pmatrix} t\,, \quad a < t < b$$

erklärten $(m + 1)$-dimensionalen reellen Raum $\mathfrak{T}$. Nunmehr haben wir $l - (m + 1) = h - 1$. Die $(m + 1)$ holomorphen Funktionen

$$K_\mu(z, w) = H_\mu(z)\ (\mu = 1, \ldots, m);\ K_{m+1}(z, w) = 0 \qquad \left(\begin{pmatrix} y \\ v \end{pmatrix} \in \mathfrak{T}\right)$$

genügen in allen $m + 1$ Variablen (z, w) Relationen der Gestalt (22). Folglich sind die Differenzengleichungen

$$G(z + e_\mu, w) - G(z, w) = H_\mu(z) \qquad \left(\mu = 1, \ldots, m;\ \begin{pmatrix} y \\ v \end{pmatrix} \in \mathfrak{T}\right), \tag{46}$$

$$G(z, w + 1) - G(z, w) = 0 \qquad \left(\begin{pmatrix} y \\ v \end{pmatrix} \in \mathfrak{T}\right)$$

nach Induktionsvoraussetzung durch eine holomorphe Funktion $G(z, w)$ lösbar. Mit Hilfe der Fourierschen Entwicklung

$$G(z, w) = \sum_{\varrho = -\infty}^{\infty} G_\varrho(z)\, e^{2\pi i \varrho w} \qquad \left(\begin{pmatrix} y \\ v \end{pmatrix} \in \mathfrak{T}\right),$$

des Eindeutigkeitssatzes für Fourierreihen und der Formeln (46) ergibt sich jetzt

$$G_0(z + e_\mu) - G_0(z) = H_\mu(z) \quad (\mu = 1, \ldots, m; z \in \mathfrak{Z}(\mathfrak{Y}))\,.$$

Beweis von Satz 2: Man schneide die Pyramide $\mathfrak{Y}$, deren Spitze im Punkte a des y-Raumes liege, mit der Einheitshyperkugel

$$\mathfrak{E}: (y - a)'(y - a) = 1\,,$$

wähle auf $\mathfrak{Y} \cap \mathfrak{E}$ eine überall dichte abzählbare Punktmenge p_ϱ derart, daß die ersten m der von a nach p_ϱ weisenden Vektoren

$$r_\varrho = p_\varrho - a \qquad (\varrho = 1, 2, \ldots)$$

linear unabhängig sind, und bilde die Matrizen

$$R^{(m,\varrho)} = (r_1, r_2, \ldots, r_\varrho) \qquad (\varrho = m, m+1, \ldots)\,.$$

Die monoton wachsende Folge der Pyramiden

$$\mathfrak{Y}_\varrho : y = R_\varrho t,\ t^{(\varrho,1)} > 0 \qquad (\varrho = m, m+1, \ldots)$$

schöpft $\mathfrak{Y}$ aus. Also ist Satz 2 auf Hilfssatz 2 und Hilfssatz 5 zurückgeführt.

§ 3. Vorbereitende Überlegungen zum Beweis von Satz 3

Auf Grund der Periodizitätsbedingungen (20) sowie der Holomorphie lassen sich die Funktionen $Q(U; Z)$ in Fourierreihen

$$Q(U; Z) = \sum_T a(U, T)\, e^{2\pi i \operatorname{Sp}(TZ)} \tag{47}$$

entwickeln; $T^{(n)}$ durchläuft alle halbganzen symmetrischen Matrizen. Die Gleichungen (21) sind wegen $\operatorname{Sp}(TZ[U]) = \operatorname{Sp}(T[U']Z)$ mit den Relationen

$$a(U_1 U_2, T) = a(U_2, T[U_1'^{-1}]) + a(U_1, T) \tag{48}$$

zwischen den Koeffizienten dieser Fourierreihen identisch. Der Forderung (18) Rechnung tragend, setzen wir auch

$$P(Z) = \sum_T b(T)\, e^{2\pi i \operatorname{Sp}(TZ)}\,, \tag{49}$$

wobei wieder über alle halbganzen symmetrischen Matrizen $T^{(n)}$ summiert wird. Um die Differenzengleichungen (19) zu lösen, müssen dann die Größen $b(T)$ so bestimmt werden, daß die Beziehungen

$$b(T[U'^{-1}]) - b(T) = a(U, T) \tag{50}$$

gelten und die Fourierreihe (49) in der verallgemeinerten oberen Halbebene $\mathfrak{Z}$ konvergiert. Offenbar genügt es, die beiden folgenden Sonderfälle zu betrachten.

Positiver Fall: $\quad a(U, T) = 0 \qquad (T \ngtr 0)$,

Nicht positiver Fall: $\quad a(U, T) = 0 \qquad (T > 0)$.

Satz 3 ergibt sich nun aus den folgenden Aussagen.

Satz 4: *Im positiven Falle sind die Relationen* (48) *notwendig und hinreichend für die Lösbarkeit des Gleichungssystems* (50) *durch geeignete Zahlen* $b(T)$ *mit*

$b(T) = 0$ $(T \not\geq 0)$. *Konvergieren alle Fourierreihen* (47) *in der verallgemeinerten oberen Halbebene* $\mathfrak{Z}$, *dann lassen sich die* $b(T)$ *so wählen, daß auch die Fourierreihe* (49) *in* $\mathfrak{Z}$ *konvergiert* $(n \geqq 2)$.

Satz 5: *Im nicht positiven Falle ist das Gleichungssystem* (50) *dann und nur dann durch Zahlen* $b(T)$, *für welche die Fourierreihe* (49) *konvergiert, lösbar, wenn die Relationen* (48) *bestehen und die Fourierreihen* (47) *in der verallgemeinerten oberen Halbebene* $\mathfrak{Z}$ *konvergieren. Die* $b(T)$ $(T \not\geq 0)$ *sind eindeutig bestimmt* $(n \geqq 3)$.

Zum Beweise dieser beiden Sätze benötigen wir einige Tatsachen, die zunächst hergeleitet werden.

§ 4. Unimodulare Matrizen

Man definiere

$$(51)\quad \left.\begin{cases} A_{\iota\varkappa}^{(n)} = (a_{\iota\varkappa;\nu\mu}) : a_{\iota\varkappa;\nu\nu} = 1\ (\nu = 1, \ldots, n),\ a_{\iota\varkappa;\iota\varkappa} = 1,\ a_{\iota\varkappa;\nu\mu} = 0 \text{ sonst} \\ B_{\iota\varkappa}^{(n)} = (b_{\iota\varkappa;\nu\mu}) : b_{\iota\varkappa;\nu\nu} = 1\ (\nu \neq \iota, \varkappa;\ \nu = 1, \ldots, n),\ b_{\iota\varkappa;\iota\varkappa} = b_{\iota\varkappa;\varkappa\iota} = 1, \\ \qquad\qquad b_{\iota\varkappa;\nu\mu} = 0 \text{ sonst} \\ C_{\iota}^{(n)} = (c_{\iota;\nu\mu}) : c_{\iota;\nu\nu} = 1\ (\nu \neq \iota;\ \nu = 1, \ldots, n),\ c_{\iota;\iota\iota} = -1,\ c_{\iota;\nu\mu} = 0 \text{ sonst} \end{cases}\right\}\ (\iota \neq \varkappa)$$

Diese Matrizen sind unimodular, und Kronecker [8] bewies

Hilfssatz 6: *Die Gruppe* Ω *der* n*-reihigen unimodularen Matrizen wird durch die Elemente*

$$A_{12}, B_{12}, \ldots, B_{1n}$$

erzeugt $(n \geqq 2)$.

Wir verstehen unter Ω_r die Gruppe aller unimodularen Matrizen der Form

$$U^{(n)} = \begin{pmatrix} E^{(r)} & 0 \\ U_1 & U_2^{(n-r)} \end{pmatrix}.$$

Der Buchstabe E bezeichnet die Einheitsmatrix.

Hilfssatz 7: *Die Elemente*

$$A_{\iota\varkappa}(\iota = r+1, \ldots, n;\ \varkappa = 1, \ldots, r);\ A_{r+1,r+2};\ B_{r+1,r+2};\ \ldots;\ B_{r+1,n}$$

erzeugen $\Omega_r (n \geqq 2)$.

Beweis: Es ist

$$U = \begin{pmatrix} E & 0 \\ U_1 & U_2 \end{pmatrix} = \begin{pmatrix} E & 0 \\ U_1 & E \end{pmatrix} \begin{pmatrix} E & 0 \\ 0 & U_2 \end{pmatrix}.$$

Wendet man (51) auf den ersten dieser Faktoren und Hilfssatz 6 auf U_2 an, so folgt die Behauptung.

Eine unimodulare Matrix U heiße *von endlicher Ordnung*, wenn $U^t = E$ für irgendeine ganze Zahl $t \neq 0$ gilt. Eine *Inversion* werde durch $U^2 = E$ erklärt. Als Beispiele für Inversionen nehme man die Matrizen $B_{\iota\varkappa}$ und C_ι.

Hilfssatz 8: *Die Gruppe* Ω_r *ist durch endlich viele Inversionen erzeugbar* $(n \geqq 1)$.

Beweis: Da die Elemente $B_{\iota\varkappa}$ Inversionen sind, braucht man nach Hilfssatz 7 nur zu zeigen, daß die Matrizen $A_{\iota\varkappa}$ $(\iota = r+1, \ldots, n;\ \varkappa = 1, \ldots, n)$ sich als Produkt von Inversionen darstellen lassen. Letzteres ergibt sich aus

$$A_{\iota\varkappa} = C_\iota(C_\iota A_{\iota\varkappa}),\ C_\iota^2 = E,\ (C_\iota A_{\iota\varkappa})^2 = E,\ C_\iota \in \Omega_r(\iota = r+1, \ldots, n;\ \varkappa = 1, \ldots, n)\,.$$

Hilfssatz 9: *Jede unimodulare Matrix $U^{(n)}$ gestattet eine Zerlegung*

$$U = P_1 W P_2$$

mit

$$W^{(n)} = \begin{pmatrix} W_1^{(2)} & 0 \\ 0 & E^{(n-2)} \end{pmatrix}.$$

Die Symbole P_1, P_2 bezeichnen gewisse Produkte der Matrizen

$$A_{\iota\varkappa}^{\pm 1}(\iota = 1, \ldots, n;\ \varkappa = 2, \ldots, n)\,,\quad B_{\iota\varkappa}(\iota, \varkappa = 2, \ldots, n) \quad (n \geqq 2)\,.$$

Beweis: Es sei

$$g = \begin{pmatrix} g_1^{(1,1)} \\ g_2^{(n-1,1)} \end{pmatrix}$$

die erste Spalte der Matrix U, d der größte gemeinsame Teiler der Elemente von g_2 und

$$d_2^{(n-1,1)} = \begin{pmatrix} d \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Dann existiert eine unimodulare Matrix $V_1^{(n-1)}$, für welche die Beziehung

$$g_2 = V_1 d_2$$

besteht, sowie eine unimodulare Matrix $W_1^{(2)}$, deren erste Spalte

$$\begin{pmatrix} g_1 \\ d \end{pmatrix}$$

ist. Setzen wir nun

$$P_1^{(n)} = \begin{pmatrix} 1 & 0 \\ 0 & V_1 \end{pmatrix},\quad W^{(n)} = \begin{pmatrix} W_1 & 0 \\ 0 & E^{(n-2)} \end{pmatrix},\quad P_2 = W^{-1} P_1^{-1} U\,,$$

so liegen beide Matrizen P_1', P_2' in Ω_1; auf Grund von Hilfssatz 7 sind P_1, P_2 Produkte der in Hilfssatz 9 angegebenen Art.

Hilfssatz 10: *Zu jeder ganzen Matrix $G^{(n)}$ mit verschwindender Determinante gibt es eine unimodulare Matrix $D^{(n)}$, so daß die Summe*

$$D + \mu G$$

bei allen ganzen Zahlen μ unimodular ist $(n \geqq 1)$.

Beweis: Wegen $\operatorname{Det} G = 0$ können wir zwei unimodulare Matrizen U, V finden, für die das Produkt UGV in und unterhalb der Hauptdiagonalen Nullen hat. Mit $D = U^{-1}V^{-1}$ wird

$$\pm \operatorname{Det}(D + \mu G) = \operatorname{Det}(E + \mu\, UGV) = 1\,.$$

Nachstehende Relationen zwischen unimodularen Matrizen dienen den späteren Betrachtungen.

(52) $$C_\iota^2 = E \qquad (n \geqq 1),$$

(53) $$B_{\iota\varkappa}^2 = E \qquad (n \geqq 2),$$

(54) $$A_{\varkappa\iota} = A_{\iota\varkappa}[B_{\iota\varkappa}] \qquad (n \geqq 2),$$

(55) $$C_\varkappa = C_\iota[B_{\iota\varkappa}] \qquad (n \geqq 2),$$

(56) $$C_\iota = A_{\varkappa\iota} A_{\iota\varkappa}^{-1} A_{\varkappa\iota} B_{\iota\varkappa} \qquad (n \geqq 2),$$

(57) $$A_{\lambda\iota} = A_{\lambda\varkappa}[B_{\iota\varkappa}] \qquad (n \geqq 3),$$

(58) $$(B_{\iota\lambda} B_{\varkappa\lambda})^3 = E \qquad (n \geqq 3),$$

(59) $$B_{\iota\varkappa}[C_\lambda] = B_{\iota\varkappa} \qquad (n \geqq 3),$$

(60) $$A_{\iota\lambda} A_{\lambda\varkappa}^{-1} A_{\iota\lambda}^{-1} A_{\lambda\varkappa} A_{\iota\varkappa} = E \qquad (n \geqq 3),$$

(61) $$A_{\lambda\iota}^\alpha A_{\lambda\varkappa}^\beta = A_{\lambda\varkappa}^\beta A_{\lambda\iota}^\alpha \qquad (n \geqq 3),$$

(62) $$A_{\iota\lambda}^\alpha A_{\varkappa\lambda}^\beta = A_{\varkappa\lambda}^\beta A_{\iota\lambda}^\alpha \qquad (n \geqq 3),$$

(63) $$A_{13}^\alpha W A_{13}^\beta = W A_{13}^\beta W^{-1} A_{13}^\alpha W \qquad (n \geqq 3).$$

In diesen Gleichungen bedeuten die Indices ι, $\varkappa$, λ, sofern sie auftreten, voneinander verschiedene der Nummern $1, \ldots, n$ und die Exponenten α, β irgendwelche ganze Zahlen. Die Matrix W hat die Gestalt

$$W^{(n)} = \begin{pmatrix} W_1^{(2)} & 0 \\ 0 & E^{(n-2)} \end{pmatrix}.$$

§ 5. *Quadratische Formen*

Im Raume $\mathfrak{R}$ der reellen symmetrischen Matrizen $S^{(n)}$ betrachte man die Bereiche

$$\mathfrak{S} : S \geqq 0\,, \quad \mathfrak{T} : S > 0\,,$$

die abgeschlossene Minkowskische Pyramide [9, 12] $\mathfrak{M}$ sowie die mit Hilfe der Ungleichungen

$$\mathrm{Sp}\, S \leqq \mathrm{Sp}(S[U]) \qquad (U \in \Omega)$$

definierte Pyramide $\mathfrak{P}$. Dabei bezeichne Ω wieder die Gruppe der unimodularen Matrizen $U^{(n)}$. Bekanntlich [9, 12] liegt die Minkowskische Pyramide $\mathfrak{M}$ in $\mathfrak{S}$, und da es zu jedem $S \ngeqq 0$ eine unimodulare Matrix U mit

$$\mathrm{Sp}(S[U]) < \mathrm{Sp}\, S$$

gibt, folgt $\mathfrak{P} \subset \mathfrak{S}$. Für die Gebiete

$$\mathfrak{N} = \mathfrak{M} \cap \mathfrak{T}\,, \quad \mathfrak{Q} = \mathfrak{P} \cap \mathfrak{T}$$

wies BAILY [2] die Existenz einer endlichen Menge $\mathfrak{x} \subset \Omega$ nach, so daß

(64) $$\mathfrak{Q} \subset \bigcup_{V \in \mathfrak{x}} \mathfrak{N}[V]$$

gilt. Nur endlich viele der Bereiche $\mathfrak{R}[U]$ $(U \in \Omega)$ grenzen an $\mathfrak{R}$ [9, 12]. Ist daher der Durchschnitt

$$\mathfrak{Q}[U] \cap \mathfrak{Q}$$

nicht leer, so muß U einer endlichen Menge $\mathfrak{u} \subset \Omega$ angehören. Ähnlich wie in der Reduktionstheorie der positiven quadratischen Formen [9, 12] für die Minkowskische Pyramide $\mathfrak{M}$ zeigt man

Hilfssatz 11: *Eine Matrix $S \in \mathfrak{R}$ liegt dann und nur dann in der Pyramide $\mathfrak{P}$, wenn die Ungleichungen*

$$\mathrm{Sp}\, S \leq \mathrm{Sp}(S[U]) \qquad (U \in \mathfrak{u})$$

gelten $(n \geq 2)$.

Hilfssatz 12: *Es gibt eine endliche Menge $\mathfrak{v} \in \Omega$ mit folgender Eigenschaft. Sind zwei Matrizen S, $T \in \mathfrak{P}$ durch eine unimodulare Matrix ineinander transformierbar, so besteht eine Beziehung*

$$T = S[U] \qquad (U \in \mathfrak{v})\ (n \geq 2)\,.$$

Beweis: Auf Grund der Konvexität der Pyramiden $\mathfrak{P}$, $\mathfrak{Q}$ ist $\mathfrak{P}$ die Abschließung von $\mathfrak{Q}$ im Raume $\mathfrak{R}$, nach (64) also

$$\mathfrak{P} \subset \bigcup_{V \in \mathfrak{s}} \mathfrak{M}[V]\,.$$

Diese Formel erlaubt es, S, $T \in \mathfrak{M}$ anzunehmen. Jetzt haben wir

$$S = \begin{pmatrix} 0 & 0 \\ 0 & S_1^{(r)} \end{pmatrix}, \quad T = \begin{pmatrix} 0 & 0 \\ 0 & T_1^{(r)} \end{pmatrix}, \quad (\mathrm{Det}\, S_1)\,(\mathrm{Det}\, T_1) \neq 0\,. \tag{65}$$

Gemäß Voraussetzung gilt

$$T = S[W]\,, \tag{66}$$

wobei W eine gewisse unimodulare Matrix bedeutet. Aus (65), (66) folgt

$$W^{(n)} = \begin{pmatrix} W_1 & W_2 \\ 0 & U_1^{(r)} \end{pmatrix}, \quad T_1 = S_1[U_1]\,.$$

Die Kästchen S_1, T_1 sind nach Minkowski reduziert. Für U_1 kommen daher nur endlich viele Möglichkeiten in Frage. Da S bei Transformation mit

$$U^{(n)} = \begin{pmatrix} E & 0 \\ 0 & U_1 \end{pmatrix}$$

in T übergeht, erkennt man die Richtigkeit von Hilfssatz 12.

Hilfssatz 13: *Für eine Matrix $S^{(n)} > 0$ und genügend kleines positives $p^{(1)}$ ist*

$$S - pE > 0 \qquad , \qquad (n \geq 1)\,.$$

Beweis: Variiert die reelle Spalte $x^{(n,1)}$ auf der Einheitshyperkugel $x'x = 1$, so besitzt die quadratische Form $S[x]$ ein positives Minimum q. Jede Zahl p des Intervalls $0 < p < q$ befriedigt die Ungleichung

$$S - pE > 0\,.$$

Hilfssatz 14: *Eine positive reelle symmetrische Matrix $S^{(n)}$ läßt sich stets in der Gestalt*

$$S = v_1 U_1 U_1' + \cdots + v_h U_h U_h' \qquad (v_\eta \geq 0;\ U_\eta^{(n)} \in \Omega;\ \eta = 1, \ldots, h)$$

schreiben. Bei rationalem S können die Parameter v_η rational gewählt werden $(n \geqq 1)$.

Beweis: Es sei $\mathfrak{U}$ die konvexe Pyramide, bestehend aus allen Matrizen

$$S = v_1 U_1 U_1' + \cdots + v_h U_h U_h' \quad (v_\eta \geqq 0;\ U_\eta \in \Omega;\ \eta = 1,\ldots,h;\ h = 1,2,3,\ldots),$$

$\mathfrak{V}$ die Abschließung von $\mathfrak{U}$ in $\mathfrak{R}$, ferner $\mathrm{Rd}\,\mathfrak{S}$ der Rand von $\mathfrak{S}$ in $\mathfrak{R}$. Wir wollen

$$\mathrm{Rd}\,\mathfrak{S} \subset \mathfrak{V} \tag{67}$$

nachweisen. Durchläuft $G^{(n)}$ alle ganzen Matrizen, deren Determinante verschwindet, und p sämtliche reellen Zahlen, dann liegen die Produkte

$$S = p^2 G G'$$

in $\mathrm{Rd}\,\mathfrak{S}$ dicht. Wegen der Abgeschlossenheit des Gebiets $\mathfrak{V}$ reicht es deshalb aus, die Formel

$$p^2 G G' \in \mathfrak{V}$$

für die Gesamtheit der oben angegebenen Matrizen pG zu zeigen. Hierzu bezeichne man mit μ einen ganzen Parameter, bestimme die Matrix D wie in Hilfssatz 10 und bilde die Ausdrücke

$$\mu^{-2} p^2 (D + \mu G)(D' + \mu G') = p^2 (G + \mu^{-1} D)(G' + \mu^{-1} D') \quad (\mu \neq 0).$$

Letztere liegen in $\mathfrak{U}$; ferner häufen sie sich gegen den Punkt $p^2 G G'$. Damit ist (67) als gültig erkannt. (67) und die Konvexität von $\mathfrak{V}$ ziehen $\mathfrak{S} \subset \mathfrak{V}$, also $\mathfrak{T} \subset \mathfrak{V}$ nach sich. Benutzen wir die Tatsachen, daß $\mathfrak{T}$ eine offene Punktmenge, $\mathfrak{V}$ aber die Abschließung der konvexen Pyramide $\mathfrak{U}$ ist, so ergibt sich $\mathfrak{T} \subset \mathfrak{U}$. Jede Matrix $S \in \mathfrak{T}$ gestattet folglich eine Summendarstellung der in Hilfssatz 14 angegebenen Art. In dieser lasse man alle Summanden mit $v_\eta = 0$ fort, führe die Maximalzahl r der linear unabhängigen unter den Matrizen $U_1 U_1', \ldots, U_h U_h'$ ein und wähle eine Indizierung, bei der die Produkte $U_1 U_1', \ldots, U_r U_r'$ nicht linear voneinander abhängen. Indem wir jetzt sämtliche Parameter ein wenig ändern, können wir $v_{r+1}, \ldots, v_h$ rational machen. Für rationales S sind dann $v_1, \ldots, v_r$ von selbst rational. Das vollendet den Beweis von Hilfssatz 14.

Hilfssatz 15: *Zu jeder rationalen symmetrischen Matrix $S^{(n)}$ vom Range r gibt es eine unimodulare Matrix $V^{(n)}$, welche S in die Gestalt*

$$S[V] = \begin{pmatrix} S_1^{(r)} & 0 \\ 0 & 0 \end{pmatrix}$$

transformiert $(n \geqq 1)$.

Beweis: Ist $\operatorname{Det} S = 0$, so existiert eine unimodulare Matrix U, deren letzte Spalte g die Gleichung $Sg = 0$ erfüllt. Mithin hat man

$$S[U] = \begin{pmatrix} S_2^{(n-1)} & 0 \\ 0 & 0 \end{pmatrix}.$$

Im Falle $\operatorname{Det} S_2 = 0$ wende man dasselbe Verfahren auf S_2 an usf.

Es bedeute $S^{(n)}$ eine reelle symmetrische Matrix. Unter der *Einheitengruppe* $\Lambda(S)$ von S verstehen wir die Gesamtheit aller unimodularen Matrizen $U^{(n)}$ mit

$$S[U] = S.$$

Hilfssatz 16: *Die symmetrische Matrix $S^{(n)}$ sei rational und nicht negativ. Dann wird die Einheitengruppe $\Lambda(S)$ durch Elemente endlicher Ordnung erzeugt* ($n \geqq 1$).

Beweis: Aus

$$T = S[V] \qquad (V \in \Omega)$$

folgt

$$V\Lambda(T)\,V^{-1} = \Lambda(S)\,.$$

Die Gruppen $\Lambda(S)$, $\Lambda(T)$ sind daher isomorph, und wegen Hilfssatz 15 können wir

$$S = \begin{pmatrix} S_1^{(r)} & 0 \\ 0 & 0 \end{pmatrix},$$

(68) $$S_1 > 0$$

voraussetzen. Jetzt zerspaltet sich jede Matrix $U \in \Lambda(S)$ in der Form

$$U = W_1 W_2,\ W_1^{(n)} = \begin{pmatrix} U_1^{(r)} & 0 \\ 0 & E \end{pmatrix},\ W_2^{(n)} \in \Omega_r\,,$$

(69) $$U_1 \in \Lambda(S_1)\,.$$

Vermöge (68), (69) besitzt U_1 endliche Ordnung [9, 12]. Hilfssatz 8 besagt, daß W_2 Produkt von Inversionen ist.

Die Vielzahl der halbganzen symmetrischen Matrizen $S^{(n)} = (s_{\varkappa\lambda})$, deren Elemente den Ungleichungen

(70) $$s_1, \ldots, s_n \geqq 0\,,$$

(71) $$2|s_{\varkappa\lambda}| \leqq s_\varkappa + s_\lambda \qquad (\varkappa \neq \lambda;\ \varkappa, \lambda = 1, \ldots, n)$$

genügen, heiße $\mathfrak{A}$. Die Menge $\mathfrak{B}$ umfasse alle halbganzen symmetrischen Matrizen, die nicht in $\mathfrak{A}$ liegen. Offenbar hat keine Matrix $S \in \mathfrak{A}$ negative Spur. Man setze $k = \frac{n(n+1)}{2}$ und bezeichne mit σ eine positive reelle Zahl. Dann liefert ein einfaches Abzählen unter Ausnutzung der Formeln (70), (71) das folgende Ergebnis.

Hilfssatz 17: *Für höchstens* $(2\sigma + 1)^k$ *verschiedene der Matrizen* $S \in \mathfrak{A}$ *ist*

$$\operatorname{Sp} S \leqq \sigma \qquad (n \geqq 1)\,.$$

Wir erklären $\mathfrak{a}$ als die Gesamtheit der in (51) definierten Matrizen $A_{\iota\varkappa}(\iota \neq \varkappa;\ \iota, \varkappa = 1, \ldots, n)$ sowie ihrer Inversen.

Hilfssatz 18: *Jedem* $S \in \mathfrak{B}$ *können wir eine Folge halbganzer Matrizen*

$$S_\nu^{(n)}(S) \in \mathfrak{B} \qquad (\nu = 0, 1, 2, \ldots)$$

zuordnen, welche nachstehende Bedingungen erfüllt.

a) $$S_0(S) = S\,,$$

b) $$S_{\nu+1}(S) = S_\nu(S)\,[W_\nu(S)] \qquad (W_\nu(S) \in \mathfrak{a};\ \nu = 0, 1, 2, \ldots)\,,$$

c) $$\operatorname{Sp} S_{\nu+1}(S) < \operatorname{Sp} S_\nu(S) \qquad (\nu = 0, 1, 2, \ldots)\,.$$

d) *Es seien σ, τ zwei ganze Zahlen mit $\sigma \leqq \tau$. Laufen S über $\mathfrak{B}$ und ν von 0 bis $+\infty$ unter den Nebenbedingungen*

$$\mathrm{Sp}\, S \leqq \tau\,, \quad \mathrm{Sp}\, S_\nu(S) = \sigma\,,$$

so sind höchstens $4n^3(\tau - \sigma + 1)$ der Matrizen $S_\nu(S)$ einander gleich ($n \geqq 2$).

Beweis: Man betrachte die Teilmenge $\mathfrak{C} \subset \mathfrak{B}$, bestehend aus allen halbganzen Matrizen, die ein negatives Diagonalelement haben. Für $S \in \mathfrak{B}$, $S \notin \mathfrak{C}$ gelten die Ungleichungen (71) nicht sämtlich; bei passender Wahl der Indices μ, ν folgt daher

$$2|s_{\mu\nu}| > s_\mu + s_\nu\,,$$

d. h. wenigstens eine der beiden Matrizen $S[A_{\mu\nu}]$, $S[A_{\mu\nu}^{-1}]$ liegt in $\mathfrak{C}$, und ihre Spur ist kleiner als $\mathrm{Sp}\, S$. Zu jeder Matrix $S \in \mathfrak{B}$, $S \notin \mathfrak{C}$ gibt es also ein Element $W(S) \in \mathfrak{a}$ mit

$$\mathrm{Sp}(S[W(S)]) < \mathrm{Sp}\, S\,, \quad S[W(S)] \in \mathfrak{C}\,.$$

Natürlich können zwei Matrizen $S[W(S)]$, $T[W(T)]$ auch dann übereinstimmen, wenn S, T verschieden sind. Die Anzahl der untereinander gleichen Matrizen $S[W(S)]$ übersteigt jedoch nicht die Anzahl $a = 2(n^2 - n)$ der Elemente von $\mathfrak{a}$. Den Matrizen S Rechnung tragend, welche von vornherein $\mathfrak{C}$ angehören, addieren wir zu diesen a Möglichkeiten noch eine weitere Möglichkeit und erhalten die Abschätzung

$$a + 1 = 2(n^2 - n) + 1 < 2n^2\,.$$

Damit ist die Untersuchung auf $S \in \mathfrak{C}$ zurückgeführt.

Für eine Matrix $S \in \mathfrak{C}$, deren erstes Diagonalelement s_1 negativ ist, definiere man

$$S_\nu(S) = S[A_{12}^{\pm\nu}] \qquad (\nu = 0, 1, 2, \ldots)\,,$$

wobei das Vorzeichen im Exponenten so bestimmt werde, daß $\mathrm{Sp}\, S_\nu(S)$ bei wachsendem ν monoton fällt. Es sei S eine feste halbganze Matrix, $s_1 < 0$. T laufe über $\mathfrak{C}$ und μ von 0 bis $+\infty$ unter den Nebenbedingungen

$$t_1 < 0\,, \quad \mathrm{Sp}\, T \leqq \tau\,, \quad \mathrm{Sp}\, S_\mu(T) = \sigma\,.$$

Aus

$$S_\mu(T) = S_\nu(S) \tag{72}$$

folgt $T = S[A_{12}^{\varrho}]$ mit einer gewissen ganzen Zahl ϱ. Ferner gilt

$$\sigma = \mathrm{Sp}\, S_\mu(T) \leqq \mathrm{Sp}\, T \leqq \tau\,.$$

Die Anzahl der Matrizen $S_\mu(T)$, die der Gleichung (72) genügen, ist also höchstens so groß wie die Anzahl b der ganzen ϱ, für welche die Ungleichung

$$\sigma \leqq \mathrm{Sp}\, S[A_{12}^{\varrho}] \leqq \tau$$

besteht. Als Funktion von ϱ betrachtet, stellt

$$\mathrm{Sp}\, S[A_{12}^{\varrho}] = s_1\varrho^2 + \cdots$$

eine Parabel dar. s_1 ist ganz und negativ, also $s_1 \leqq -1$. Daraus ergibt sich

$$b \leqq 2(\tau - \sigma + 1).$$

Um den Fall $s_1, \ldots, s_{\lambda-1} \geqq 0$; $s_\lambda < 0$ $(\lambda > 1)$ zu erledigen, setzen wir

$$S_\nu(S) = S[A_{\lambda 1}^{\pm\nu}],$$

machen bei festem λ die vorigen Überlegungen und bekommen wieder die Abschätzung $2(\tau - \sigma + 1)$. Im ganzen benötigen wir n Matrizen aus $\mathfrak{a}$, nämlich

$$A_{12}, A_{21}, \ldots, A_{n1}.$$

Dieses gibt einen Faktor n. Hinzu kommen die früher genannten $2n^2$, so daß höchstens

$$4n^3(\tau - \sigma + 1)$$

der Matrizen $S_\nu(S)$ unter den in d) genannten Voraussetzungen einander gleich sind.

§ 6. *Fourierreihen*

In diesem Paragraphen wird die Konvergenz gewisser Fourierreihen untersucht.

Hilfssatz 19: *Dann und nur dann konvergiert die Fourierreihe*

$$\sum_{T \geqq 0} d(T)\, e^{2\pi i\, \mathrm{Sp}(TZ)}$$

für $Z \in \mathfrak{Z}$, wenn es zu jeder positiven reellen Zahl p eine Konstante $c_1(p) > 0$ gibt, derart daß die Abschätzungen

$$|d(T)| \leqq c_1(p)\, e^{p\, \mathrm{Sp}\, T} \tag{73}$$

bestehen; $T^{(n)}$ durchläuft sämtliche nicht negativen halbganzen symmetrischen Matrizen $(n \geqq 1)$.

Beweis: Die in Hilfssatz 19 genannte Fourierreihe ist genau dann konvergent, falls sie absolut konvergiert, d. h.

$$\sum_{T \geqq 0} |d(T)|\, e^{-2\pi i\, \mathrm{Sp}(TY)} \qquad (Y > 0) \tag{74}$$

einen Grenzwert besitzt. Auf Grund von Hilfssatz 13 gilt bei hinreichend kleinem positiven p die Ungleichung

$$2\pi Y - pE > 0.$$

Folglich besteht die notwendige und hinreichende Bedingung für die Konvergenz der Reihe (74) darin, daß

$$\sum_{T \geqq 0} |d(T)| e^{-p\, \mathrm{Sp}\, T} \tag{75}$$

einen Limes $l(p)$ hat. p bedeutet hierbei eine beliebige positive reelle Zahl. Die Konvergenz von (75) zieht (73) mit $c_1(p) = l(p)$ nach sich. Nehmen wir umgekehrt die Richtigkeit von (73) an. Jetzt wird (75) durch

$$c_1\left(\frac{p}{2}\right) \sum_{T \geqq 0} e^{-\frac{p}{2}\, \mathrm{Sp}\, T}$$

und dieses vermöge $T \in \mathfrak{A}$ sowie Hilfssatz 17 von der konvergenten Reihe

$$c_1\left(\frac{p}{2}\right) \sum_{\sigma=0}^{\infty} (2\sigma+1)^k e^{-\frac{p}{2}\sigma}$$

majorisiert. Somit konvergiert (75).

Hilfssatz 20: *Die Fourierreihe*

$$\sum_{T \geqq 0} d(T) e^{2\pi i \operatorname{Sp}(TZ)}$$

konvergiert dann und nur dann für $Z \in \mathfrak{Z}$, *falls es zu jeder unimodularen Matrix* $U^{(n)}$ *und positiven reellen Zahl* p *eine Konstante* $c_2(U, p) > 0$ *gibt, so daß die Ungleichungen*

$$\sum_{\operatorname{Sp}(T[U]) = \mu} |d(T)| \leqq c_2(U, p)\, e^{p\mu} \quad (\mu = 0, \pm 1, \pm 2, \ldots) \tag{76}$$

gelten. Zu summieren hat man über die halbganzen Matrizen $T \geqq 0 (n \geqq 1)$.

Beweis: Unsere Fourierreihe ist genau dann konvergent, wenn sie absolut konvergiert, d. h.

$$\sum_{T \geqq 0} |d(T)|\, e^{2\pi \operatorname{Sp}(TY)} \tag{77}$$

für $Y \in \mathfrak{T}$ einen Grenzwert hat. Die Gesamtheit der Matrizen

$$Y = \frac{q}{2\pi} R, \tag{78}$$

mit einer reellen Zahl $q > 0$ und rationalen symmetrischen Matrix $R^{(n)} > 0$, liegt im Raume $\mathfrak{T}$ dicht. Daher kann man die Konvergenzuntersuchungen auf diese Matrizen (78) beschränken. Wir tragen (78) in (77) ein und bekommen die Summe

$$\sum_{\lambda} \left(\sum_{\substack{\operatorname{Sp}(TR) = \lambda \\ T \geqq 0}} |d(T)|\, e^{-q\lambda} \right), \tag{79}$$

erstreckt über alle rationalen Zahlen λ, welche, mit dem Hauptnenner der Elemente von R multipliziert, eine ganze Zahl ergeben. Hat die Reihe (79) einen Limes $l(R, q)$, so besteht die Ungleichung

$$\sum_{\substack{\operatorname{Sp}(TR) = \lambda \\ T \geqq 0}} |d(T)| \leqq c_3(R, q)\, e^{q\lambda} \tag{80}$$

mit der Konstanten $c_3(R, q) = l(R, q)$. Nehmen wir umgekehrt an, daß die Abschätzungen (80) für gewisse positive Werte $c_3(R, q)$ sowie die früher erklärten R, q, λ erfüllt sind. Dann ist der konvergente Ausdruck

$$c_3\left(R, \frac{q}{2}\right) \sum_{\lambda \geqq 0} e^{-\frac{q}{2}\lambda} + c_3(R, 2q) \sum_{\lambda < 0} e^{q\lambda}$$

Majorante der Summe (79), letztere demnach konvergent. Die Ungleichungen (80) erweisen sich als notwendig und hinreichend für die Konvergenz der betrachteten Fourierreihe.

Die Abschätzungen (80) gelten bei $R = R_1 + R_2$ sowie allen q, λ, falls sie für R_1 und R_2 richtig sind. Aus $\operatorname{Sp}(TR) = \lambda$ folgt nämlich $\operatorname{Sp}(TR_1) \leq \frac{\lambda}{2}$ oder $\operatorname{Sp}(TR_2) \leq \frac{\lambda}{2}$, also

$$\sum_{\substack{\operatorname{Sp}(TR)=\lambda \\ T \geqq 0}} |d(T)| \leq \sum_{\varkappa \leq \frac{\lambda}{2}} \Bigg(\sum_{\substack{\operatorname{Sp}(TR_1)=\varkappa \\ T \geqq 0}} |d(T)| + \sum_{\substack{\operatorname{Sp}(TR_2)=\varkappa \\ T \geqq 0}} |d(T)| \Bigg)$$

$$\leq (c_3(R_1, 2q) + c_3(R_2, 2q)) \sum_{\varkappa \leq \frac{\lambda}{2}} e^{2q\varkappa} = c_3(R, q)\, e^{q\lambda}$$

mit

$$c_3(R, q) = (c_3(R_1, 2q) + c_3(R_2, 2q)) \sum_{\varrho \leq 0} e^{q\varrho}\,;$$

die hierbei auftretenden rationalen Summationsindices λ, $\varkappa$, ϱ besitzen beschränkte Nenner. Aus dieser Überlegung und Hilfssatz 14 folgt das Bestehen der Ungleichungen (80) für alle rationalen Matrizen $R > 0$, falls sie bei sämtlichen Matrizen

$$R = vUU' \tag{81}$$

gelten. U durchläuft dabei Ω, v die positiven rationalen Zahlen. (81) in (80) eingesetzt, liefert mittels der Substitutionen

$$\mu = \frac{\lambda}{v},\; p = qv,\; c_2(U, p) = c_3(vUU', q)$$

die Ungleichungen (76).

§ 7. *Beweis von Satz 4*

Die Lösbarkeit des Gleichungssystems (50) zieht

$$a(U, T) = 0 \qquad (U' \in \Lambda(T)) \tag{82}$$

nach sich. Zum Beweise von Satz 4 leiten wir jetzt diese Formeln (82) aus den Relationen (48) her. $U_1 = U_2 = E$ in (48) eingetragen, ergibt $a(E, T) = 0$. Für ein Element endlicher Ordnung U mit

$$U^t = E,\quad t \neq 0,\quad U' \in \Lambda(T)$$

bekommen wir

$$a(U, T) = t^{-1} a(E, T) = 0$$

und wegen Hilfssatz 16 und (48) die Behauptung.

Man nenne zwei symmetrische Matrizen äquivalent, falls sie sich durch eine unimodulare Matrix ineinander transformieren lassen, wähle aus jeder Klasse äquivalenter nicht negativer halbganzer Matrizen einen Repräsentanten S und führe die Zahlen

$$b(S[V'^{-1}]) = a(V, S) \qquad (V \in \Omega) \tag{83}$$

ein; diese sind auf Grund der Beziehungen (82) wohldefiniert. Es sei T eine beliebige halbganze symmetrische Matrix, $T = S[V'^{-1}]$, ferner $U \in \Omega$. Aus (83), (48) folgt

$$b(T[U'^{-1}]) - b(T) = a(VU, S) - a(V, S) = a(U, T)\,.$$

Damit haben wir das Gleichungssystem (50) gelöst.

Hilfssatz 21: *Liegen sämtliche Repräsentanten S in der früher erklärten Pyramide $\mathfrak{P}$, so konvergiert die Fourierreihe*

$$\sum_{T \geq 0} b(T)\, e^{2\pi i \operatorname{Sp}(TZ)}$$

für $Z \in \mathfrak{Z}$ $(n \geq 2)$.

Beweis: Gehört die Matrix $T = T_0$ nicht zu $\mathfrak{P}$, dann gibt es nach Hilfssatz 11 eine unimodulare Matrix $U_0 \in \mathfrak{u}$ mit

$$T_1 = T_0[U_0], \ \operatorname{Sp} T_1 < \operatorname{Sp} T_0 .$$

Im Falle $T_1 \notin \mathfrak{P}$ bestimme man ein $U_1 \in \mathfrak{u}$, derart daß

$$T_2 = T_1[U_1], \ \operatorname{Sp} T_2 < \operatorname{Sp} T_1$$

gilt. So fortfahrend erhält man eine Folge halbganzer symmetrischer Matrizen $T_1, T_2, T_3, \ldots$, welche den Bedingungen

$$T_\eta = T_{\eta-1}[U_{\eta-1}],\ U_{\eta-1} \in \mathfrak{u} \qquad (\eta = 1, 2, \ldots),$$

$$\operatorname{Sp} T_\eta < \operatorname{Sp} T_{\eta-1} \qquad (\eta = 1, 2, \ldots) \tag{84}$$

genügen. Das Verfahren bricht ab; denn da die Spur einer halbganzen Matrix ganz ist, bekommen wir vermöge der Ungleichungen (84) nach höchstens $\operatorname{Sp} T$ Schritten eine Matrix $T_h \in \mathfrak{P}$. Hilfssatz 12 lehrt die Existenz einer unimodularen Matrix U_h mit

$$T_{h+1} = T_h[U_h] = S \qquad (U_h \in \mathfrak{v}).$$

Nun wird

$$T = S[U_h^{-1} U_{h-1}^{-1} \cdots U_0^{-1}],$$

gemäß (83), (48) also

$$b(T) = \sum_{\eta=0}^{h} a(U'_\eta, T_{\eta+1}). \tag{85}$$

Aus der Konvergenz der Fourierreihen (47), Hilfssatz 19 und (84) schließen wir die Abschätzungen

$$|a(U'_\eta, T_{\eta+1})| \leq c_1\left(\frac{p}{2}\right) e^{\frac{p}{2}\operatorname{Sp} T_{\eta+1}} \leq c_1\left(\frac{p}{2}\right) e^{\frac{p}{2}\operatorname{Sp} T} \qquad (\eta = 0, 1, \ldots, h).$$

Hierbei konnte $c_1\left(\frac{p}{2}\right)$ von η unabhängig gewählt werden, weil die Matrizen U_η in der endlichen Menge $\mathfrak{u} \cup \mathfrak{v}$ liegen. Somit haben wir wegen (85) die Ungleichungen

$$|b(T)| \leq c_1\left(\frac{p}{2}\right)(1 + h)\, e^{\frac{p}{2}\operatorname{Sp} T} \leq c_1\left(\frac{p}{2}\right)(1 + \operatorname{Sp} T)\, e^{\frac{p}{2}\operatorname{Sp} T} \leq c_4(p)\, e^{p \operatorname{Sp} T}$$

für eine passende Konstante $c_4(p)$. Durch nochmalige Benutzung von Hilfssatz 19 folgt die Behauptung.

§ 8. *Beweis von Satz 5*

Gegeben seien eine halbganze symmetrische Matrix $T^{(n)} \geqq 0$ und eine unimodulare Matrix $U^{(n)}$. Mit $\mathfrak{p}(U, T)$ bezeichne man die Gesamtheit der Folgen (V_ι, T_ι), welche den nachstehenden Bedingungen genügen.

e) $$T_0 = T\,,$$

f) $$T_{\iota+1} = T_\iota[V_\iota] \qquad (V_\iota \in \Omega;\ \iota = 0, 1, 2, \ldots),$$

g) $$\mathrm{Sp}(T_{\iota+1}[U]) < \mathrm{Sp}(T_\iota[U]) \quad \textit{für fast alle } \iota = 0, 1, 2, \ldots.$$

Für fast alle kürzen wir im folgenden gelegentlich durch *f. a.* ab.

Nehmen wir an, Satz 5 sei richtig. Dann befriedigen die Koeffizienten $b(T)$ die Abschätzungen (76), woraus sich

$$|b(T)| \leqq c_2(U, p)\, e^{p\,\mathrm{Sp}(T[U])} \qquad (U \in \Omega) \tag{86}$$

ergibt. Die Formeln (50), (86) ziehen für jede Folge $(V_\iota, T_\iota) \in \mathfrak{p}(U, T)$ den Zusammenhang

$$b(T) = -\sum_{\iota=0}^{\infty} a(V_\iota'^{-1}, T_\iota) \tag{87}$$

nach sich, d. h. unsere Differenzengleichungen (50) besitzen höchstens ein System von Lösungen $b(T)$, bei dem die Fourierreihe (49) konvergiert. Zum Beweise des Satzes 5 leiten wir jetzt die Beziehungen (87) mit geeigneten Koeffizienten $b(T)$ aus (48) und der Konvergenz der Reihen (47) her.

Hilfssatz 22: *Die Summen*

$$-\sum_{\iota=0}^{\infty} a(V_\iota'^{-1}, T_\iota) \qquad ((V_\iota, T_\iota) \in \mathfrak{p}(U, T)) \tag{88}$$

konvergieren. Ihr Wert hängt nicht von der Folge (V_ι, T_ι), *sondern nur von der Matrix* T *ab; wir nennen ihn* $b(T)$. *Diese Größen* $b(T)$ *lösen die Gleichungen* (50) *und genügen den Abschätzungen* (86) $(n \geqq 3)$.

Beweis: Aus der Konvergenz der Fourierreihe (47) erhalten wir unter Benutzung von Hilfssatz 20 die Ungleichungen

$$|a(W, T)| \leqq c_5(W;\, U, p)\, e^{p\,\mathrm{Sp}(T[U])} \qquad (W;\, U \in \Omega), \tag{89}$$

insbesondere also

$$|a(W, T)| \leqq c_6(W;\, p)\, e^{p\,\mathrm{Sp}\,T} \qquad (W \in \Omega) \tag{90}$$

mit geeigneten positiven Zahlen c_5, c_6. Es bedeute $\mathfrak{q}(T)$ die Gesamtheit der Folgen $(V_\iota, T_\iota) \in \mathfrak{p}(E, T)$, bei denen für fast alle Indizes $\iota = 0, 1, 2, \ldots$ die ersten Diagonalelemente der Matrizen T_ι negativ und die V_ι gleich einer Bestimmten der Matrizen $A_{1\nu}^{\varepsilon}$ $(\nu = 2, \ldots, n;\ \varepsilon = \pm 1)$ sind. Wählt man $(V_\iota, T_\iota) \in \mathfrak{q}(T)$, so besitzt die Summe (88) wegen (90) einen Grenzwert; wir werden zeigen, daß dieser nur von T abhängt.

Gegeben seien zwei Folgen (V_ι, T_ι), $(U_\iota, S_\iota) \in \mathfrak{q}(T)$. Man bestimme eine Zahl r, für welche die ersten Diagonalelemente der Matrizen T_ι, S_ι $(\iota \geqq r)$ negativ und die Gleichungen

$$V_\iota = A_{1\nu}^{\varepsilon_1},\ U_\iota = A_{1\mu}^{\varepsilon_2} \qquad (\nu, \mu = 2, \ldots, n;\ \varepsilon_1, \varepsilon_2 = \pm 1)$$

erfüllt sind. Wir definieren

$$U = V_{r-1}^{-1} V_{r-2}^{-1} \ldots V_1^{-1} V_0^{-1} U_0 U_1 \ldots U_{r-2} U_{r-1},$$

bekommen

$$S_r = T_r[U]$$

und schreiben die Differenz

$$d = \sum_{\iota=0}^{\infty} a(V_\iota'^{-1}, T_\iota) - \sum_{\iota=0}^{\infty} a(U_\iota'^{-1}, S_\iota)$$

in der Form

$$d = \sum_{\iota=0}^{\infty} a(A_{1\nu}'^{-\iota\varepsilon_1}, T_r[A_{1\nu}^{\iota\varepsilon_1}]) - a(U'^{-1}, T_r) - \sum_{\iota=0}^{\infty} a(A_{1\mu}'^{-\iota\varepsilon_2}, S_r[A_{1\mu}^{\iota\varepsilon_2}]) .$$

Unser Problem können wir also wie folgt formulieren. Vorgelegt seien zwei halbganze symmetrische Matrizen S, T mit negativen ersten Diagonalelementen; ferner gelte $S = T[U]$ $(U \in \Omega)$. Zu beweisen ist das Verschwinden der Ausdrücke

$$(91) \qquad d = \sum_{\iota=0}^{\infty} a(A_{1\nu}'^{-\iota\varepsilon_1}, T[A_{1\nu}^{\iota\varepsilon_1}]) - a(U'^{-1}, T) - \sum_{\iota=0}^{\infty} a(A_{1\mu}'^{-\iota\varepsilon_2}, S[A_{1\mu}^{\iota\varepsilon_2}])$$

$$(\nu, \mu = 2, \ldots, n;\ \varepsilon_1, \varepsilon_2 = \pm 1) .$$

Dazu machen wir nachstehende Überlegung. Vermöge (48) gestattet die rechte Seite von (91) die Umformung

$$d = \sum_{\iota=\varrho}^{\infty} a(A_{1\nu}'^{-\iota\varepsilon_1}, T[A_{1\nu}^{\iota\varepsilon_1}]) - a(A_{1\nu}'^{\varrho\varepsilon_1} U'^{-1} A_{1\mu}'^{-\varrho\varepsilon_2}, T[A_{1\nu}^{\varrho\varepsilon_1}]) -$$

$$- \sum_{\iota=\varrho}^{\infty} a(A_{1\mu}'^{-\iota\varepsilon_2}, S[A_{1\mu}^{\iota\varepsilon_2}]) \qquad (\varrho = 0, 1, 2, \ldots) .$$

Nehmen wir nun an, es existiere eine Relation

$$(92) \quad A_{1\nu}^{-\varrho\varepsilon_1} U A_{1\mu}^{\varrho\varepsilon_2} = W_{\varrho 1} \ldots W_{\varrho\alpha(\varrho)} \quad (W_{\varrho\beta} \in \Omega;\ \beta = 1, \ldots, \alpha(\varrho);\ \varrho = 0, 1, 2, \ldots)$$

mit diesen Eigenschaften:

h) *Die unimodularen Matrizen $W_{\varrho\beta}$ gehören einer endlichen, von ϱ unabhängigen Menge an.*

i) $$\alpha(\varrho) \leq c_7 \varrho \qquad (\textit{f. a. } \varrho = 0, 1, 2, \ldots) .$$

j) *Die durch*

$$R_{\varrho 1} = T[A_{1\nu}^{\varrho\varepsilon_1}];\ R_{\varrho,\beta+1} = R_{\varrho\beta}[W_{\varrho\beta}] \quad (\beta = 1, \ldots, \alpha(\varrho) - 1)$$

erklärten Matrizen $R_{\varrho\beta}$ genügen den Ungleichungen

$$\operatorname{Sp} R_{\varrho\beta} \leq -c_8^{-1} \varrho \qquad (\beta = 1, \ldots, \alpha(\varrho);\ \textit{f. a. } \varrho = 0, 1, 2, \ldots) .$$

Hierbei hängen die Konstanten c_7, c_8 nicht von ϱ ab.

Dann ergibt sich aus den Bedingungen (48) der Zusammenhang

$$d = \sum_{\iota=\varrho}^{\infty} a(A_{1\nu}'^{-\iota\varepsilon_1}, T[A_{1\nu}^{\iota\varepsilon_1}]) - \sum_{\beta=1}^{\alpha(\varrho)} a(W_{\varrho\beta}'^{-1}, R_{\varrho\beta}) - \sum_{\iota=\varrho}^{\infty} a(A_{1\mu}'^{-\iota\varepsilon_2}, S[A_{1\mu}^{\iota\varepsilon_2}])$$

$$(\varrho = 0, 1, 2, \ldots)$$

und hieraus wegen (90) sowie den Forderungen h) bis j) die Abschätzung

$$|d| \leq c_8 \varrho e^{-\varrho} \qquad (f.\,a.\ \varrho = 0, 1, 2, \ldots),$$

also $d = 0$. Mithin ist das Problem darauf zurückgeführt, geeignete Relationen der Gruppe Ω zu finden.

Wir untersuchen als erstes den Fall $U = E$; hierzu werde $d(\nu, \varepsilon_1; \mu, \varepsilon_2)$ statt d geschrieben. Für $\nu \neq \mu$ stellt (61) mit $\alpha = -\varrho\varepsilon_1$, $\beta = \varrho\varepsilon_2$ eine Relation der gewünschten Art dar; somit folgt $d(\nu, \varepsilon_1; \mu, \varepsilon_2) = 0$. Bei $\nu = \mu$ bestimme man einen Index $\lambda \neq 1, \nu$. Es wird

$$d(\nu, \varepsilon_1; \nu, \varepsilon_2) = d(\nu, \varepsilon_1; \lambda, 1) + d(\lambda, 1; \nu, \varepsilon_2) = 0\,.$$

Nun sei U beliebig unimodular. Auf Grund der eben durchgeführten Betrachtungen hängen die Werte der beiden in (91) auftretenden Summen nicht von den Zahlen $\nu, \varepsilon_1; \mu, \varepsilon_2$ ab. Wir wählen daher die Bezeichnung $d(U; T, S)$. Eine Zerlegung

$$U = U_1 U_2;\ U_1, U_2 \in \Omega;\ T[U_1] = R;\ R[U_2] = S$$

zieht vermöge (48) die Gleichung

$$d(U; T, S) = d(U_1; T, R) + d(U_2; R, S)$$

nach sich, falls das erste Diagonalelement der Matrix R negativ ist. Darum braucht man das Verschwinden der Differenzen $d(U; T, S)$ nur für die in Hilfssatz 9 angegebenen Matrizen $U = W, A_{\iota\varkappa}^{\pm 1}, B_{\iota\varkappa}$ zu beweisen.

$U = W$: Man setze $\nu = \mu = 3$, $\varepsilon_1 = 1$, benutze die Beziehung (63) mit $\alpha = -\varrho$, $\beta = \varrho\varepsilon_2$ und lege ε_2 so fest, daß die Forderung j) erfüllt ist.

$U = A_{\iota\varkappa}^{\pm 1}, B_{\iota\varkappa}$ $(\iota, \varkappa = 2, \ldots, n)$: Wir wenden (62), (57) an.

$U = A_{1\mu}^{\pm 1}$ $(\mu = 2, \ldots, n)$: $d(A_{1\mu}^{\pm 1}; T, S) = d(E; T, T) = 0$.

Für alle Folgen $(V_\iota, T_\iota) \in \mathfrak{q}(T)$ ist also der Wert der Summen (88) eine Funktion $b(T)$ von T allein.

Es sei $U \in \Omega$, ferner die Folge

$$(V_0, T_0); (V_1, T_1); (V_2, T_2); \ldots$$

in $\mathfrak{q}(T[U'^{-1}])$ gelegen. Dann gehört die Folge

$$(U'^{-1}, T); (V_0, T_0); (V_1, T_1); (V_2, T_2); \ldots$$

zur Klasse $\mathfrak{q}(T)$; nach (87) gilt

$$b(T) = -a(U, T) + b(T[U'^{-1}])\,,$$

d. h. die Größen $b(T)$ lösen die Differenzengleichungen (50).

Zu fest vorgegebener unimodularer Matrix U wollen wir eine Folge $(V_\iota, T_\iota) \in \mathfrak{q}(T)$ konstruieren, derart daß die Matrizen V_ι in einer nur von U abhängigen endlichen Menge $\mathfrak{w}(U)$ enthalten und die Formeln

$$\text{(93)} \qquad \mathrm{Sp}(T_{\iota+1}[U]) < \mathrm{Sp}(T_\iota[U]) \qquad (\iota = 0, 1, 2, \ldots)$$

erfüllt sind. Die Matrix $S_0 = T[U]$ liegt nicht in $\mathfrak{S}$, also erst recht nicht in $\mathfrak{P}$. Wegen Hilfssatz 11 gibt es daher eine Matrix $W_0 \in \mathfrak{u}$, für welche die Ungleichung

$$\mathrm{Sp}(S_0[W_0]) < \mathrm{Sp}\, S_0$$

besteht. Wir wenden dieses Verfahren wiederholt an und bekommen eine Kette von Matrizenpaaren (W_ι, S_ι) mit

$$S_{\iota+1} = S_\iota[W_\iota],\ W_\iota \in \mathfrak{u},\ \mathrm{Sp}\, S_{\iota+1} < \mathrm{Sp}\, S_\iota\ (\iota = 1, \ldots, \alpha - 1),\ \mathrm{Sp}\, S_\alpha < 0\,.$$

Die Matrix S_α besitzt ein negatives Diagonalelement s_ϱ. Im Falle $\varrho \neq 2$ betrachte man die Folge

$$S_{\iota+1} = S_\iota[W_\iota],\ W_\iota = A_{\varrho 2}^{\varepsilon_1} \qquad (\iota = \alpha, \ldots, \beta - 1, \varepsilon_1 = \pm 1)\,.$$

Bei richtiger Wahl des Exponenten ε_1 fallen die Spuren $\mathrm{Sp}\, S_\iota$ stark monoton; für genügend großes β ist das zweite Diagonalelement der Matrix S_β negativ. Man definiere

$$S_{\iota+1} = S_\iota[W_\iota],\ W_\iota = A_{21}^{\varepsilon_2} \qquad (\iota = \beta, \ldots, \gamma - 1; \varepsilon_2 = \pm 1)\,.$$

Wieder nehmen die Größen $\mathrm{Sp}\, S_\iota$ bei passendem ε_2 mit wachsendem ι ab. $S_\gamma[U^{-1}]$ hat für geeignetes γ negatives erstes Diagonalelement. Wir setzen

$$V_\iota = \begin{Bmatrix} U W_\iota U^{-1} (\iota = 0, \ldots, \gamma - 1) \\ A_{12}^{\varepsilon_3} (\iota = \gamma, \gamma + 1, \ldots; \varepsilon_3 = \pm 1) \end{Bmatrix},\ T_\iota = \begin{Bmatrix} S_\iota[U^{-1}] \ (\iota = 0, 1, \ldots, \gamma) \\ T_{\iota-1}[A_{12}^{\varepsilon_3}] \ (\iota = \gamma + 1, \gamma + 2, \ldots) \end{Bmatrix},$$

$$\mathfrak{w}(U) = U \mathfrak{u} U^{-1} \cup U \mathfrak{a} U^{-1} \cup \mathfrak{a}\,.$$

Die gewünschten Ungleichungen (93) erhält man, wenn ε_3 richtig bestimmt wird. Jetzt ergeben sich die Abschätzungen (86) ohne Schwierigkeiten. Da alle unimodularen Matrizen V_ι der endlichen Menge $\mathfrak{w}(U)$ angehören, können wir die in (89) auftretenden Konstanten $c_5(V_\iota'^{-1}; U, p)$ durch eine gemeinsame Schranke $c_{10}(U, p)$ majorisieren und bekommen aus (87), (93)

$$|b(T)| \leq c_{10}(U, p) \sum_{\iota=0}^{\infty} e^{p\, \mathrm{Sp}(T_\iota[U])} \leq c_2(U, p)\, e^{p\, \mathrm{Sp}(T[U])}$$

mit

$$c_2(U, p) = c_{10}(U, p) \left(\sum_{\iota=0}^{\infty} e^{-p\iota} \right).$$

Vermöge der Formeln (50), (86) gelten die Gleichungen (87) für sämtliche Folgen $(V_\iota, T_\iota) \in \mathfrak{p}(U, T)$. Der Beweis von Hilfssatz 22 ist beendet.

Es bleibt übrig, die Konvergenz der Fourierreihe

$$\sum_{T \geqq 0} b(T)\, e^{2\pi i\, \mathrm{Sp}(TZ)}$$

zu zeigen. Wir wählen eine feste Matrix $U \in \Omega$, ferner eine ganze Zahl μ, nehmen die Zerspaltung

$$(94) \qquad \sum_{\substack{\mathrm{Sp}(T[U]) = \mu \\ T \geqq 0}} |b(T)| = \sum_{\substack{\mathrm{Sp}(T[U]) = \mu \\ T[U] \in \mathfrak{A},\, T \geqq 0}} |b(T)| + \sum_{\substack{\mathrm{Sp}(T[U]) = \mu \\ T[U] \in \mathfrak{B}}} |b(T)|$$

vor und schätzen beide Teile getrennt ab. Auf die erste Summe wende man die Ungleichungen (86) sowie Hilfssatz 17 an. Dann ergibt sich

$$\sum_{\substack{\mathrm{Sp}(T[U]) = \mu \\ T[U] \in \mathfrak{A},\, T \geqq 0}} |b(T)| \leq c_2\left(U, \frac{p}{2}\right) (2\mu + 1)^k e^{\frac{p}{2}\mu} \leq c_{11}(U, p)\, e^{p\mu}$$

für eine geeignete Konstante $c_{11}(U, p)$. Zur Abschätzung des zweiten Ausdrucks der rechten Seite von (94) benutzen wir Hilfssatz 18, setzen darin $S = T[U]$ und führen die Folgen

$$(V_\nu(T), T_\nu(T)) = (U W_\nu(T[U]) U^{-1}, S_\nu(T[U]) [U^{-1}]) \in \mathfrak{p}(U, T)$$

ein. Es bestehen die Ungleichungen

$$\operatorname{Sp}(T_{\nu+1}[U]) < \operatorname{Sp}(T_\nu[U]) \qquad (\nu = 0, 1, 2, \ldots);$$

die Matrizen $V_\nu(T)$ liegen in der endlichen Menge $U\mathfrak{a}U^{-1}$. Laufen $T[U]$ über $\mathfrak{B}$ sowie ν von 0 bis $+\infty$ unter den Nebenbedingungen

$$\operatorname{Sp}(T[U]) \leq \tau, \quad \operatorname{Sp}(T_\nu(T)[U]) = \sigma,$$

so sind höchstens $4n^3(\tau - \sigma + 1)$ der Matrizen $T_\nu(T)$ einander gleich. Nunmehr folgt

$$\sum_{\substack{\operatorname{Sp}(T[U]) = \mu \\ T[U] \in \mathfrak{S}}} |b(T)| \leq \sum_{\substack{\operatorname{Sp}(T[U]) = \mu \\ T[U] \in \mathfrak{S}}} \left\{ \sum_{\nu = 0}^{\infty} |a(V_\nu'^{-1}(T), T_\nu(T))| \right\} \leq$$

$$\leq 4n^3 \sum_{V \in U\mathfrak{a}U^{-1}} \left\{ \sum_{\iota = -\infty}^{\mu} (\mu - \iota + 1) \left(\sum_{\operatorname{Sp}(S[U]) = \iota} |a(V'^{-1}, S)| \right) \right\}.$$

Die Konvergenz der Reihen (47) zieht gemäß Hilfssatz 20

$$\sum_{\operatorname{Sp}(S[U]) = \iota} |a(V'^{-1}, S)| \leq c_{12}(V; U, p)\, e^{\varrho\iota}$$

nach sich; die Symbole $c_{12}(V; U, p)$ bedeuten wieder positive Konstanten. Also wird

$$\sum_{\substack{\operatorname{Sp}(T[U]) = \mu \\ T \gtrless 0}} |b(T)| \leq c_{13}(U, p)\, e^{\varrho\mu}$$

mit

$$c_{13}(U, p) = c_{11}(U, p) + 4n^3 \left(\sum_{V \in U\mathfrak{a}U^{-1}} c_{12}(V; U, p) \right) \left(\sum_{\iota = 0}^{\infty} (\iota + 1)\, e^{-\varrho\iota} \right).$$

Auf Grund von Hilfssatz 20 ist Satz 5 bewiesen.

§ 9. Ein weiteres Differenzengleichungsproblem

Jedem Element $U^{(n)} \in \Omega$ sei eine Matrix $M^{(n)}(U)$ zugeordnet. Wir wollen untersuchen, für welche Systeme von Matrizen $M(U)$ eine halbganze symmetrische Matrix $D^{(n)}$ die Differenzengleichungen

$$D[U'] - D = M(U) \tag{95}$$

befriedigt. Bei Bestehen der Beziehung (95) sind die Matrizen $M(U)$ halbganz und symmetrisch; ferner gilt

$$M(U_1 U_2) = M(U_2)[U_1'] + M(U_1). \tag{96}$$

Satz 6: ***Dann und nur dann ist das Differenzengleichungssystem* (95) *durch eine halbganze symmetrische Matrix D lösbar, wenn die Matrizen* $M(U)$ *halbganz, symmetrisch und die Relationen* (96) *erfüllt sind. Die Gleichungen* (95) *bestimmen D eindeutig* ($n \geq 3$).**

Beweis: Man definiere

$$L(U) = M(U) - D[U'] + D\,.$$

Die symmetrische halbganze Matrix D läßt sich auf genau eine Weise so wählen, daß die ersten sowie $(n-\iota)$ letzten Elemente der ersten Zeilen und Spalten der Matrizen $L(B_{1\iota})$ $(\iota = 2, \ldots, n)$ verschwinden und $L(A_{12})$ die Form

$$(97) \qquad L(A_{12}) = \begin{pmatrix} 0 & l & 0 \\ l & l_1 & l_2 \\ 0 & l_2' & L_1^{(n-2)} \end{pmatrix} \qquad \left(l = 0, \tfrac{1}{2}\right)$$

besitzt. Zu zeigen ist $L(U) = 0$ $(U \in \Omega)$. Auf Grund von (96) genügen die Matrizen $L(U)$ den Bedingungen

$$(98) \qquad L(U_1 U_2) = L(U_2)\,[U_1'] + L(U_1)\,;$$

daraus folgt $L(E) = 0$ sowie

$$(99) \qquad L(U^{-1}) = -L(U)\,[U'^{-1}]\,.$$

Die Relationen (53) liefern

$$L(B_{1\iota})\,[B_{1\iota}] + L(B_{1\iota}) = 0\,,$$

d. h.

$$(100) \qquad L(B_{1\iota}) = \begin{pmatrix} 0 & k_{\iota 2} \;\cdots\; k_{\iota\,\iota-1} & 0 & \\ k_{\iota 2} & & -k_{\iota 2} & \\ \vdots & 0 & \vdots & 0 \\ k_{\iota\,\iota-1} & & -k_{\iota\,\iota-1} & \\ 0 & -k_{\iota 2} \cdots -k_{\iota\,\iota-1} & 0 & \\ & 0 & & 0 \end{pmatrix} (\iota = 2, \ldots, n)\,.$$

Insbesondere gilt $L(B_{12}) = 0$, und wegen (54), (98) erhalten wir

$$L(A_{21}) = L(A_{12})\,[B_{12}]\,.$$

Mit Hilfe der Formeln (56), (97) bis (99) bekommt man hieraus

$$L(C_2) = \begin{pmatrix} l_1 & l_1 + l & l_2 \\ l_1 + l & l_1 & 3 l_2 \\ l_2' & 3 l_2' & L_1 \end{pmatrix}.$$

Der Zusammenhang (52) lehrt, daß die Größen l_1, l_2, L_1 Null sind. Das zweite Element der ersten Zeile von $L(B_{1\iota}[C_2])$ $(\iota = 3, \ldots, n)$ hat den Wert $l - k_{\iota 2}$. Die Beziehungen (59), (100) ergeben $l = 2k_{\iota 2}$, mithin ist l ganz, daher $l = k_{\iota 2} = 0$, $L(C_2) = 0$. Es sei r eine ganze Zahl, $3 \leq r \leq n$. Wir machen die Induktionsannahme

$$k_{\iota\varrho} = 0 \qquad (\varrho < \iota;\ \iota = 3, \ldots, n;\ \varrho = 2, \ldots, r-1).$$

Gleichung (100) zieht dann $L(B_{1\varrho}) = 0$ $(\varrho = 2, \ldots, r)$ nach sich. Indem wir C_r mittels (55), (58) durch B_{12}, B_{1r}, C_2 ausdrücken, folgt $L(C_r) = 0$, vermöge (59) also

$$k_{\iota r} = 0 \qquad (\iota = r + 1, \ldots, n)\,.$$

Jetzt wissen wir, daß die Matrizen

$$L(A_{12}),\ L(B_{12}), \ldots, L(B_{1n})$$

verschwinden. Die Relationen (98) und Hilfssatz 6 vollenden den Beweis.

Kapitel III. Ergebnisse

§ 1. *Klassifikation*

Die Ausdrücke

$$g(S,T) = J(E,S;Z+T) + J(E,T;Z) - J(E,T;Z+S) - J(E,S;Z) \tag{101}$$

sind auf Grund der Relationen (10) kongruent Null, d. h. ganze Zahlen. Diese bleiben invariant, wenn man das Exponentensystem $J(U, S; Z)$ durch ein äquivalentes ersetzt. Es bestehen die Beziehungen

$$g(S, T) = -g(T, S)\,, \tag{102}$$

$$g(S_1+S_2, T) = g(S_1, T) + g(S_2, T),\; g(S, T_1+T_2) = g(S, T_1) + g(S, T_2)\,, \tag{103}$$

welche sich allein aus der Definition (101) sowie den Formeln (10), angewandt für die Gruppe der Translationen

$$Z^* = Z + S\,,$$

ergeben. Jetzt benutzen wir, daß die Funktionen $J(U, S; Z)$ ein Exponentensystem zur Gruppe der ganzen Modulsubstitutionen

$$Z^* = Z[U] + S$$

bilden. Aus (10) sowie der Identität

$$(Z + S)[U] = Z[U] + S[U] \tag{104}$$

leiten wir die Kongruenz

$$J(E, S[U]; Z[U]) - J(E, S; Z) \equiv J(U, 0; Z + S) - J(U, 0; Z)$$

und hieraus die Gleichung

$$\begin{aligned} &J(E, S[U]; (Z + T)[U]) - J(E, S; Z + T) - J(E, S[U]; Z[U]) + J(E, S; Z) \\ &= J(U, 0; Z + S + T) - J(U, 0; Z + T) - J(U, 0; Z + S) + J(U, 0; Z) \end{aligned}$$

her. Ihre rechte Seite ändert sich bei Vertauschung der Matrizen S, T nicht; daraus folgt

$$g(S[U], T[U]) = g(S, T) \qquad (U \in \Omega)\,. \tag{105}$$

Die Zusammenhänge (102), (103), (105) ziehen nun das Verschwinden sämtlicher Zahlen $g(S, T)$ nach sich. Zum Beweise beachte man, daß sich jede ganze symmetrische Matrix $S^{(n)}$ als Linearkombination der Matrizen

$$\begin{cases} S_\iota = (s_{\iota;\nu\mu}) : s_{\iota;\iota\iota} = 1,\ s_{\iota;\nu\mu} = 0 \text{ sonst} & (\iota = 1, 2, \ldots, n)\,, \\ S_{\iota\varkappa} = (s_{\iota\varkappa;\nu\mu}) : s_{\iota\varkappa;\iota\varkappa} = s_{\iota\varkappa;\varkappa\iota} = 1,\ s_{\iota\varkappa;\nu\mu} = 0 \text{ sonst} & (\iota \neq \varkappa;\ \iota, \varkappa = 1, \ldots, n) \end{cases} \tag{106}$$

mit ganzen Koeffizienten schreiben läßt. Wegen (103) genügt es daher, $g(S, T) = 0$ für alle Matrizen S, T der in (106) angegebenen Gestalt zu zeigen. Als erstes nehmen wir $S = S_\iota$, $T = S_\varkappa$ bzw. $S = S_{\iota\varkappa}$, $T = S_{\mu\nu}$ an. Dann gilt

$$S = T[B]\,, \quad T = S[B] \qquad (B \in \Omega)\,,$$

wobei B entweder eine oder das Produkt zweier der in (51) definierten Matrizen $B_{\iota\varkappa}$ ist. Aus (105) ergibt sich $g(S, T) = g(T, S)$, also $g(S, T) = 0$ ver-

möge (102). Übrig bleibt der Fall $S = S_\iota$, $T = S_{\nu\mu}$. Auf Grund von $S_{\nu\mu} = S_{\mu\nu}$, $\nu \neq \mu$ dürfen wir $\nu \neq \iota$ voraussetzen. Es folgt

$$g(S_\iota, S_\nu) = g(S_\iota[A_{\nu\mu}], S_\nu[A_{\nu\mu}]) = g(S_\iota, S_\nu + S_\mu + S_{\nu\mu}) = g(S_\iota, S_\nu) + g(S_\iota, S_\mu) + \\ + g(S_\iota, S_{\nu\mu})$$

und somit nach dem Vorhergehenden $g(S_\iota, S_{\nu\mu}) = 0$.

Weil die linken Seiten der Gleichungen (101) verschwinden, kann man die Exponenten $J(E, S; Z)$ derart durch kongruente ersetzen, daß die Relationen

$$J(E, S_1 + S_2; Z) = J(E, S_2; Z + S_1) + J(E, S_1; Z)$$

erfüllt sind. Satz 2 lehrt die Existenz eines zu $J(U, S; Z)$ äquivalenten Exponentensystems $J_1(U, S; Z)$ mit

$$J_1(E, S; Z) \equiv 0. \tag{107}$$

Aus (10), (104), (107) errechnet sich

$$J_1(U, S; Z) \equiv J_1(U, 0; Z),$$
$$J_1(U, 0; Z + S) \equiv J_1(U, 0; Z). \tag{108}$$

Man betrachte ein zu $J_1(U, S; Z)$ äquivalentes Exponentensystem. Dieses genügt genau dann den Bedingungen (107), wenn für die in (6) erklärte Funktion $L(Z)$ die Kongruenzen

$$L(Z + S) \equiv L(Z) \tag{109}$$

bestehen. Wir fassen die Zahlen

$$d_{\iota\varkappa} = \begin{cases} L(Z + S_\iota) - L(Z) & (\iota = \varkappa) \\ \frac{1}{2}(L(Z + S_{\iota\varkappa}) - L(Z)) & (\iota \neq \varkappa) \end{cases} \qquad (\iota, \varkappa = 1, \ldots, n)$$

zu einer halbganzen symmetrischen Matrix $D^{(n)} = (d_{\iota\varkappa})$ zusammen und bekommen wegen (109) die Zerspaltung

$$L(Z) = \mathrm{Sp}(DZ) + P(Z), \tag{110}$$

wobei $P(Z)$ die Perioden S besitzt. Es sei $D^{(n)}$ eine beliebige halbganze symmetrische Matrix, $P(Z)$ irgendeine in $\mathfrak{Z}$ holomorphe Funktion mit den Perioden S, ferner L durch (110) erklärt. Dann gilt (109). Die Kongruenz (108) berücksichtigend, stelle man auch jeden der Exponenten $J_1(U, 0; Z)$ als Summe einer linearen sowie einer periodischen Funktion dar, entwickle letztere in eine Fourierreihe und nenne deren konstantes Glied $j(U)$. Man erhält

$$J_1(U, 0; Z) = \mathrm{Sp}(M(U)Z) + Q(U; Z) + j(U).$$

Hierbei bedeuten die Symbole $M^{(n)}(U)$ halbganze Matrizen, $Q(U; Z)$ Fourierreihen ohne konstantes Glied, $j(U)$ komplexe Zahlen ($U \in \Omega$). Die Relationen (10) liefern

$$Q(U_1 U_2; Z) = Q(U_2; Z[U_1]) + Q(U_1; Z), \; M(U_1 U_2) = M(U_2)[U_1'] + M(U_1),$$
$$j(U_1 U_2) \equiv j(U_1) + j(U_2). \tag{111}$$

Nunmehr erkennen wir mit Hilfe der Sätze 3, 6 für $n \neq 2$ die Äquivalenz der Exponentensysteme $J_1(U, S; Z)$ und

$$J_2(U, S; Z) \equiv j(U). \tag{112}$$

Zwei verschiedene Exponentensysteme der Gestalt (112) sind nicht äquivalent. Damit ist die Klassifikation der Exponentensysteme darauf zurückgeführt, die Gesamtheit der Systeme von Zahlen $j(U)$, welche den Bedingungen (111) genügen, anzugeben.

Aus (111) schließen wir $j(E) \equiv 0$, $j(V) \equiv 0$, $\frac{1}{2}$ bei Inversionen V, also $j(U) \equiv 0, \frac{1}{2}$ $(U \in \Omega)$ vermöge Hilfssatz 8. Zwei Matrizen $A_{\varrho\sigma}$, $A_{\mu\nu}$ sind durch Produkte der in (51) definierten Matrizen $B_{\iota\varkappa}$ ineinander transformierbar. Es folgt $j(A_{\varrho\sigma}) \equiv j(A_{\mu\nu})$ und auf Grund von (60) die Kongruenz $j(A_{12}) \equiv 0$. Die Formeln (58) ziehen $j(B_{1\iota}) \equiv j(B_{1n})$, d. h. $j(B_{1\iota}) \equiv j$ $(\iota = 2, \ldots, n)$ mit einer von ι unabhängigen Zahl $j = 0$ oder $\frac{1}{2}$ nach sich. Hilfssatz 6 besagt, daß sich jede unimodulare Matrix $U^{(n)}$ als Produkt der Matrizen A_{12}, $B_{12}, \ldots, B_{1n}$ darstellen läßt. Für $\operatorname{Det} U = 1$ ist die Anzahl der Faktoren $B_{12}, \ldots, B_{1n}$ in diesem Produkt gerade; dann wird $j(U) \equiv 0$. Im Falle $\operatorname{Det} U = -1$ dagegen treten die Matrizen $B_{12}, \ldots, B_{1n}$ in ungerader Anzahl auf; wir bekommen $j(U) \equiv j$. Die Substitution $Z^* = Z[U]$ bleibt ungeändert, wenn wir $-U$ statt U einsetzen. Es folgt $j(U) = j(-U)$ und daraus $j \equiv 0$ bei ungeradem n. Mittels des Zusammenhanges (8) gehe man von den Exponentensystemen zu den Multiplikatorensystemen über. Unsere Untersuchungen ergeben das nachstehende Resultat.

Satz 7: *Für ungerade n gibt es genau eine Klasse äquivalenter Multiplikatorensysteme; diese wird durch das Multiplikatorensystem*

$$I(U, S; Z) = 1$$

repräsentiert. Bei geradem n haben wir zwei Äquivalenzklassen, denen die Repräsentanten

$$I(U, S; Z) = 1, \operatorname{Det} U$$

entsprechen $(n \neq 2)$.

In § 1 des ersten Kapitels betonten wir, daß jedes Multiplikatorensystem zu einer Form gehört. Es ist daher noch die Existenz von Formen mit den in Satz 7 genannten Multiplikatorensystemen nachzuweisen. Eine Form, zu welcher eines der in Satz 7 angegebenen Multiplikatorensysteme gehört, heiße *reduziert*.

§ 2. Reduzierte Formen

Aus [7] schließen wir

Satz 8: *Jede reduzierte Form F(Z) läßt sich in eine Fourierreihe*

$$F(Z) = \sum_{T \geq 0} f(T)\, e^{2\pi i \operatorname{Sp}(TZ)}$$

entwickeln. T durchläuft sämtliche nicht negativen halbganzen symmetrischen Matrizen $(n \geq 3)$.

Wie in § 7 des zweiten Kapitels nenne man zwei symmetrische Matrizen äquivalent, wenn sie durch eine unimodulare Matrix ineinander transformierbar sind, wähle aus jeder Klasse äquivalenter nicht negativer halbganzer symmetrischer Matrizen einen Repräsentanten S und bezeichne mit $\mathfrak{h}$ die Gesamtheit dieser Repräsentanten. $\mathfrak{k}$ bedeute die Teilmenge derjenigen Matrizen $S \in \mathfrak{h}$,

für die alle Elemente U der Einheitengruppe $\Lambda(S)$ die Determinante 1 besitzen. $\Lambda(S)$ enthält stets die Matrizen E, $-E$. Bei ungeradem n ist $\mathfrak{k}$ also leer. Es sei n gerade. Dann gilt $\operatorname{Det}(\pm E) = 1$. Liegt nun eine zu S äquivalente Matrix im Inneren der Minkowskischen Pyramide $\mathfrak{M}$, so besteht die Einheitsgruppe $\Lambda(S)$ nur aus den Elementen E, $-E$; es folgt $S \in \mathfrak{k}$. Wir führen die Funktionen

$$H(S; Z) = \sum_{U \bmod \Lambda(S)} e^{2\pi i \operatorname{Sp}(S[U]Z)} \qquad (S \in \mathfrak{h}),$$

$$K(S; Z) = \sum_{U \bmod \Lambda(S)} (\operatorname{Det} U)\, e^{2\pi i \operatorname{Sp}(S[U]Z)} \qquad (S \in \mathfrak{k})$$

ein. Summiert wird über ein volles System von Repräsentanten U der verschiedenen Linksklassen [7] der Gruppe Ω modulo $\Lambda(S)$. Eine triviale Rechnung ergibt

Satz 9: *Die Gesamtheit der reduzierten Formen stimmt mit der Schar sämtlicher in der verallgemeinerten oberen Halbebene $\mathfrak{Z}$ konvergenten Summen*

$$H(Z) = \sum_{s \in \mathfrak{h}} h(S)\, H(S; Z)$$

bzw.

$$K(Z) = \sum_{s \in \mathfrak{k}} k(S)\, K(S; Z)$$

überein $(n \geqq 3)$.

§ 3. *Automorphe Funktionen*

Es sei $\mathfrak{Z}$ eine k-dimensionale komplexe analytische Mannigfaltigkeit und Δ eine Gruppe eineindeutiger holomorpher Abbildungen von $\mathfrak{Z}$ auf sich. Unter einer *automorphen Funktion zur Gruppe* Δ verstehen wir eine in $\mathfrak{Z}$ nicht identisch verschwindende meromorphe Funktion $G(Z)$, welche den Bedingungen

$$G(\Theta Z) = G(Z) \qquad (\Theta \in \Delta)$$

genügt. Jetzt bedeute $\mathfrak{Z}$ ein einfach zusammenhängendes Gebiet im k-dimensionalen komplexen euklidischen Raum. Auf Grund von [10] läßt sich jede in $\mathfrak{Z}$ meromorphe Funktion $G(Z)$ als Quotient zweier lokal teilerfremder in $\mathfrak{Z}$ holomorpher Funktionen $F_\nu(Z)$ $(\nu = 1, 2)$ schreiben. Diese Funktionen $F_\nu(Z)$ sind nicht eindeutig bestimmt. Multipliziert man sie nämlich beide mit derselben Einheit $J(Z)$, so bleiben sie lokal teilerfremd, und ihr Quotient behält seinen Wert $G(Z)$. Uns interessiert der Fall, daß $G(Z)$ eine automorphe Funktion zur Gruppe Δ ist. Die Funktionen $F_\nu(Z)$ stellen dann Formen zur Gruppe Δ dar, zu welchen das gleiche Multiplikatorensystem gehört. Ersetzung der Formen $F_\nu(Z)$ $(\nu = 1, 2)$ durch $J(Z)\, F_\nu(Z)$ $(\nu = 1, 2)$ liefert ein äquivalentes Multiplikatorensystem; alle äquivalenten Multiplikatorensysteme sind auf diese Weise erhältlich. $J(Z)$ bedeutet dabei eine Einheit. Wir lassen $\mathfrak{Z}$ wieder die verallgemeinerte obere Halbebene sowie Δ die ganze Modulgruppe n-ten Grades sein. Es folgt

Satz 10: *Jede automorphe Funktion zur ganzen Modulgruppe n-ten Grades Δ ist Quotient zweier lokal teilerfremder reduzierter Formen* $(n \geqq 3)$.

Man verstehe unter Γ die Δ umfassende Gruppe sämtlicher Modulsubstitutionen n-ten Grades [11]

$$Z^* = (AZ + B)(CZ + D)^{-1} .$$

Aus den Sätzen 8 und 10 sowie [3] erhalten wir für $n \geqq 3$ das Bailysche [2] Ergebnis, daß jede automorphe Funktion zu Γ als Quotient zweier Modulformen n-ten Grades geschrieben werden kann.

Literatur

[1] Appell, M. P.: Sur les fonctions périodiques de deux variables. J. de. Math. Sér. IV, 7, 157—219 (1891). — [2] Baily, W. L.: Satake's Compactification of V_n^*. Amer. J. Math. 80, 348—364 (1958). — [3] Christian, U.: Zur Theorie der Modulfunktionen n-ten Grades. Math. Ann. 133, 281—297 (1957). — [4] Conforto, F.: Abelsche Funktionen und algebraische Geometrie. Springer-Verlag 1956. — [5] Guichard, M. C.: Sur la résolution de l'équation aux différences finies $G(x + 1) - G(x) = H(x)$. Ann. Ec. Norm. Sup. Sér. III, 4, 361—380 (1887). — [6] Hurwitz, A.: Sur l'intégrale finie d'une fonction entière. Ac. Math. 20, 285—312 (1897). — [7] Koecher, M.: Zur Theorie der Modulformen n-ten Grades. I. Math. Z. 59, 399—416 (1954). — [8] Kronecker, L.: Über bilineare Formen. J. reine angew. Math. 68, 273—285 (1868). — [9] Minkowski, H.: Diskontinuitätsbereich für arithmetische Äquivalenz. J. reine angew. Math. 129, 220—274 (1905). — [10] Oka, K.: Sur les fonctions analytiques de plusieurs variables. III — Deuxième problème de Cousin. J. Sci. Hiroshima Univ. Ser. A, 9, 7—19 (1939). — [11] Siegel, C. L.: Einführung in die Theorie der Modulfunktionen n-ten Grades. Math. Ann. 116, 617—657 (1939). — [12] Siegel, C. L.: Einheiten quadratischer Formen. Abh. Math. Sem. Hansischen Univ. 13, 209—239 (1940).

(Eingegangen am 21. April 1959)

BAUER, H.
Math. Annalen 138, 398—427 (1959)

Konservative Abbildungen lokal-kompakter Räume*

Von

HEINZ BAUER in Hamburg*

Einleitung

In dieser Arbeit untersuchen wir eine Klasse maßtheoretisch interessanter, stetiger Abbildungen lokal-kompakter Räume. Genauer handelt es sich um folgendes: Gegeben sei eine stetige Abbildung $\varphi : X \to Y$ eines lokal-kompakten Raumes X in einen lokal-kompakten Raum Y. Dann ordnet φ gewissen positiven (Radonschen) Maßen μ auf X ein positives Maß $\varphi(\mu)$ auf Y als Bild zu; $\varphi(\mu)$ ist definiert durch die für alle stetigen, reellen Funktionen f auf Y mit kompaktem Träger gültige Gleichung

$$\int f d\varphi(\mu) = \int f \circ \varphi \, d\mu$$

unter der Voraussetzung, daß jede der Funktionen $f \circ \varphi$ μ-integrierbar ist. Stets wird hierbei das Maß $\varphi(\mu)$ von der Menge $\varphi(X)$ getragen. Hat nun die Abbildung φ die spezielle Eigenschaft, daß umgekehrt zu jedem Maß $\nu \geqq 0$ auf Y, welches von der Menge $\varphi(X)$ getragen wird, mindestens ein Maß $\mu \geqq 0$ auf X existiert mit $\varphi(\mu) = \nu$, gehen also gewissermaßen bei Anwendung von φ auf die „zulässigen" positiven Maße auf X keine auf $\varphi(X)$ konzentrierten, positiven Maße auf Y verloren, so nennen wir die Abbildung φ *konservativ*. Es soll hier gezeigt werden, daß eine Reihe topologisch interessanter Abbildungen konservativ, also auch von Interesse für die Maßtheorie sind. Die hier gewonnenen Resultate gestatten Anwendungen in der Potentialtheorie und in der Theorie der Flächenmaße.

Die Arbeit gliedert sich in sechs Paragraphen. In § 1 stellen wir zunächst die für unsere Bedürfnisse wichtigsten Begriffe und Resultate aus der Integrationstheorie zusammen und ergänzen sie sodann durch auf unser eigentliches Ziel abgestimmte Zusätze. Hierunter fallen u. a. die ersten elementaren Eigenschaften der konservativen Abbildungen, welche dort zunächst etwas allgemeiner als soeben, nämlich ohne Voraussetzung der Stetigkeit, definiert werden.

In § 2 wird als wichtiges Beweishilfsmittel ein *Lokalisationsprinzip* bereitgestellt. Es zeigt sich nämlich, daß die φ-Darstellbarkeit eines Maßes $\nu \geqq 0$ auf Y, d. h. die Existenz eines Maßes $\mu \geqq 0$ auf X mit $\varphi(\mu) = \nu$ eine lokale

*) Der Verfasser ist der Vereinsbank in Hamburg für Forschungsmittel, welche die Fertigstellung der vorliegenden Arbeit ermöglichten, sehr dankbar.

Eigenschaft ist. Grob gesagt ist für die φ-Darstellbarkeit von ν die „φ-Darstellbarkeit im Kleinen" notwendig und hinreichend.

Eine erste Klasse topologisch interessanter Abbildungen, die sämtlich konservativ sind, wird dann in § 3 aufgezeigt. Es sind dies die *eigentlichen* und *lokal-eigentlichen* Abbildungen lokal-kompakter Räume, deren Bedeutung man z. B. aus der Theorie der holomorphen Abbildungen komplexer Räume kennt (vgl. K. Stein [18] und R. Remmert [17]). Entscheidendes Beweishilfsmittel ist hierbei neben dem Lokalisationsprinzip ein von I. Namioka [14] und Verf. [3] unabhängig und etwa gleichzeitig angegebener Fortsetzungssatz für positive Linearformen. Ferner wird der rein topologische Satz bewiesen, daß eine stetige Abbildung $\varphi : X \to Y$ eines lokal-kompakten Raumes X in einen lokal-kompakten Raum Y stets zu einer eigentlichen Abbildung $\tilde{\varphi} : \tilde{X} \to Y$ eines lokal-kompakten Raumes $\tilde{X}$ in Y und damit zu einer konservativen Abbildung fortgesetzt werden kann. Jede stetige Abbildung ist daher wenigstens die Restriktion eine konservativen Abbildung.

Unter Verwendung der in § 3 gewonnenen Ergebnisse wird in § 4 u. a. gezeigt, daß jede stetige Abbildung $\varphi : X \to Y$ eines *im Unendlichen abzählbaren* lokal-kompakten Raumes X in einen lokal-kompakten Raum Y konservativ ist.

In § 5 wird bewiesen, daß eine stetige Abbildung $\varphi : X \to Y$ stets dann konservativ ist, wenn sie einen *lokalen Schnitt* besitzt, d. h. wenn in mindestens einer Umgebung V eines jeden Punktes $y \in \varphi(X)$ eine stetige Abbildung $\sigma : V \to X$ definiert ist, für welche $\varphi \circ \sigma$ die identische Abbildung ist. Diese Eigenschaft besitzen die Projektionsabbildungen einer umfangreichen Klasse von lokal-kompakten Faserräumen.

Der abschließende § 6 beschäftigt sich mit der *Faktorisierung* konservativer Abbildungen. Zunächst wird gezeigt, daß das Transitivitätsgesetz in seiner allgemeinen Form nicht gilt, daß also die lokal-kompakten Räume als Objekte und die konservativen Abbildungen als Morphismen *keine* Kategorie bilden. Wohl aber gilt ein abgeschwächtes Transitivitätsgesetz: das Produkt $\psi \circ \varphi$ zweier konservativer Abbildungen $\varphi : X \to Y$ und $\psi : Y \to Z$ ist selbst wieder konservativ, sofern $\varphi(X) = Y$ ist. Bemerkenswerterweise gilt hiervon eine Umkehrung: ist $\psi \circ \varphi$ konservativ und $\varphi(X) = Y$, so ist wenigstens die Abbildung ψ (nicht notwendig jedoch φ) konservativ. Eine analoge Eigenschaft besitzen die holomorphen Abbildungen komplexer Räume (vgl. K. Stein [18], Satz 2; R. Remmert [17], S. 363). Gerade diese letzte Eigenschaft gestattet es, eine Theorie *konservativer Basen und Zerlegungen* zu entwickeln, die analog zur Theorie analytischer Basen und Zerlegungen von K. Stein [18] ist. Auf den Zusammenhang dieses Teils der Arbeit mit einem Faktorisierungssatz von G. T. Whyburn [19], p. 141, wird in [4] genauer eingegangen.

Es sei noch bemerkt, daß die unter [2] zitierte C. R.-Note einen Vorläufer dieser Arbeit darstellt. Ein Teil der hier bewiesenen Ergebnisse findet sich dort ohne Angabe von Beweisen und meist unter entbehrlichen, zusätzlichen Voraussetzungen.

§ 1. Darstellbare Maße und konservative Abbildungen

In diesem einleitenden Paragraphen sollen zunächst die wichtigsten Begriffe und Ergebnisse aus der Integrationstheorie Radonscher Maße zusammengestellt werden, die im folgenden benötigt werden. Wir benützen hierbei weitgehend die Terminologie von N. BOURBAKI [5], [6]; dort sind auch nähere Einzelheiten zu den folgenden Nummern 1—5 zu finden. Sodann definieren wir die für die weiteren Untersuchungen zentralen Begriffe „darstellbares Maß" und „konservative Abbildung". An einem Beispiel wird die Existenz nicht-konservativer Abbildungen nachgewiesen.

1.1. *Radonsche Maße.* — Es sei T ein lokal-kompakter[1]) topologischer Raum. Dann bezeichnen wir mit $\mathcal{K}(T)$ die Menge aller auf T stetigen, reellen Funktionen mit kompaktem Träger. $\mathcal{K}(T)$ ist hinsichtlich der üblichen Operationen und der natürlichen Ordnungsrelation $\leq$ ein geordneter Vektorraum über dem Körper $\mathbf{R}$ der reellen Zahlen. $\mathcal{K}_+(T)$ bezeichne den konvexen Kegel aller Funktionen $f \geqq 0$ aus $\mathcal{K}(T)$.

Positives (Radonsches) *Maß* auf T heißt jede positive Linearform μ auf $\mathcal{K}(T)$. Beispielsweise ist $f \to f(t)$ für jeden Punkt $t \in T$ eine solche Linearform, also ein positives Maß auf T. Dieses wird mit ε_t bezeichnet und das Maß der *Einheitsmasse im Punkte t* genannt.

Die Menge aller positiven Maße auf T bezeichnen wir mit $\mathcal{M}(T)$. Bezüglich der üblichen Operationen ist $\mathcal{M}(T)$ ein konvexer Kegel und als solcher in natürlicher Weise geordnet: $\mu \leq \nu$ ist gleichbedeutend mit $\mu(f) \leq \nu(f)$ für alle $f \in \mathcal{K}_+(T)$. Ferner trägt $\mathcal{M}(T)$ die sogenannte *vage Topologie*; sie ist nichts anderes als die Topologie der punktweisen Konvergenz auf $\mathcal{K}(T)$.

1.2. *Integrierbare und meßbare reelle Funktionen.* — Es sei μ ein positives Maß auf einem lokal-kompakten Raum T. Dann wird jeder numerischen[2]) Funktion $f \geqq 0$ auf T wie folgt das Oberintegral $\int^* f\,d\mu$ zugeordnet: Für jede nach unten halbstetige Funktion $f \geqq 0$ ist

$$\int^* f\,d\mu = \sup_{h \leq f,\, h \in \mathcal{K}_+(T)} \mu(h)\,. \tag{1.1a}$$

Für ein beliebiges $f \geqq 0$ ist

$$\int^* f\,d\mu = \inf_{g \in \mathcal{I}_+(f)} \int^* g\,d\mu\,, \tag{1.1b}$$

wenn hierbei $\mathcal{I}_+(f)$ die Menge aller nach unten halbstetigen, numerischen Funktionen g auf T mit $g \geqq f$ bezeichnet. Für jede numerische Funktion f auf T wird

$$N_1^\mu(f) = \int^* |f|\,d\mu \tag{1.2}$$

gesetzt. Offenbar ist $N_1^\mu \geqq 0$.

---

[1]) Wir verwenden den Kompaktheitsbegriff von N. BOURBAKI [7]. Lokal-kompakte Räume sind demnach auch stets Hausdorffsch (separiert).

[2]) Unter einer *numerischen* bzw. *reellen* Funktion f auf T verstehen wir eine Abbildung $f: T \to \overline{\mathbf{R}}$ bzw. $f: T \to \mathbf{R}$ von T in die durch die Adjunktion von $\pm\infty$ kompaktifizierte Zahlengerade $\overline{\mathbf{R}}$ bzw. in $\mathbf{R}$ selbst.

Die Menge $\mathscr{F}^1(\mu)$ aller *reellen* Funktionen f auf T mit $N_1^\mu(f) < +\infty$ ist bezüglich der üblichen Operationen ein Vektorraum und N_1^μ ist auf $\mathscr{F}^1(\mu)$ eine Halbnorm; N_1^μ definiert somit in $\mathscr{F}^1(\mu)$ eine lokal-konvexe Topologie: die Topologie der *Konvergenz im Mittel* (bezüglich μ). Es ist $\mathscr{K}(T) \subset \mathscr{F}^1(\mu)$ und μ stetig bezüglich dieser Topologie. Daher kann μ auf genau eine Weise fortgesetzt werden zu einer stetigen Linearform μ_0 auf der abgeschlossenen Hülle $\mathscr{L}^1(\mu)$ von $\mathscr{K}(T)$ in $\mathscr{F}^1(\mu)$. Eine reelle *Funktion* f auf T wird (μ-)*integrierbar* genannt, wenn sie in $\mathscr{L}^1(\mu)$ liegt; $\mu_0(f)$ wird das μ-*Integral* von f genannt und mit $\int f d\mu$ bezeichnet. Speziell ist also $\mu(f) = \mu_0(f) = \int f d\mu$ für jedes $f \in \mathscr{K}(T)$. Daher schreibt man für Funktionen $f \in \mathscr{K}(T)$ anstelle von $\mu(f)$ meist gleich von vornherein $\int f d\mu$. Für Funktionen $f \geqq 0$ aus $\mathscr{L}^1(\mu)$ ist

$$\int f d\mu = \int^* f d\mu = N_1^\mu(f) . \tag{1.3}$$

Daher ist $f \to \int f d\mu$ eine positive Linearform auf dem durch die natürliche Ordnungsrelation $\leqq$ geordneten Vektorraum $\mathscr{L}^1(\mu)$.

Für eine Teilmenge A von T bezeichnen wir mit χ_A stets deren charakteristische Funktion (bezüglich T). Die *Menge* A wird (μ-)*integrierbar* genannt, wenn $\chi_A \in \mathscr{L}^1(\mu)$ ist; $\mu(A) = \int \chi_A d\mu$ heißt dann das Maß von A. Die Menge A wird (μ-)*meßbar* genannt, wenn für jede in T kompakte Menge K der Durchschnitt $A \cap K$ integrierbar ist. Integrierbar sind z. B. alle kompakten und alle relativ-kompakten offenen Mengen; meßbar sind z. B. alle abgeschlossenen und alle offenen Mengen. Das System aller meßbaren Mengen ist eine Boolesche σ-Mengenalgebra mit der Einheit T.

Eine Menge $A \subset T$ heißt (μ-)*vernachlässigbar* bzw. (μ-)*lokal-vernachlässigbar* wenn A integrierbar und $\mu(A) = 0$ bzw. wenn A meßbar und $\mu(A \cap K) = 0$ ist für jede in T kompakte Menge K.

Eine numerische *Funktion* f auf T heißt (μ-)*meßbar*, wenn jede der Mengen $\{x : f(x) \geqq \alpha\}$, $\alpha \in \mathbf{R}$, meßbar ist. Eine reelle Funktion f auf T ist dann und nur dann integrierbar, wenn sie meßbar und $N_1^\mu(f) < +\infty$ ist.

1.3. *Im wesentlichen integrierbare Funktionen.* — Eine reelle Funktion f auf T heißt *im wesentlichen* (μ-)*integrierbar*, wenn sie lokal-fast-überall gleich einer integrierbaren Funktion ist, d. h. wenn eine Funktion $f' \in \mathscr{L}^1(\mu)$ existiert, für welche $f(x) = f'(x)$ ist für alle $x \in T$ mit Ausnahme der Punkte x einer lokal-vernachlässigbaren Menge. Integral einer solchen Funktion f wird dann die Zahl $\int f' d\mu$ genannt, die, wie sich zeigt, unabhängig von der speziellen Wahl von f' ist. Offenbar ist jede integrierbare Funktion auch im wesentlichen integrierbar und ihr Integral im neuen Sinne gleich dem im alten Sinne. Da somit keine Mißverständnisse entstehen können, wird auch für jede im wesentlichen integrierbare Funktion f das Integral mit $\int f d\mu$ bezeichnet.

Die Menge $\overline{\mathscr{L}}^1(\mu)$ aller im wesentlichen integrierbaren Funktionen ist hinsichtlich der üblichen Operationen und der natürlichen Ordnungsrelation $\leqq$ abermals ein geordneter Vektorraum; es ist $\mathscr{L}^1(\mu) \subset \overline{\mathscr{L}}^1(\mu)$ und $f \to \int f d\mu$ eine positive Linearform auf $\overline{\mathscr{L}}^1(\mu)$.

Der Raum $\overline{\mathscr{L}}^1(\mu)$ kann aus $\mathscr{K}(T)$ auf analoge Weise gewonnen werden

wie $\mathscr{L}^1(\mu)$ aus $\mathscr{K}(T)$. Man hat hierbei nur die Halbnorm N_1^μ durch eine Halbnorm $\overline{N}_1^\mu \leq N_1^\mu$ zu ersetzen, die wie folgt definiert ist: Für jede numerische Funktion f auf T ist

$$\overline{N}_1^\mu(f) = \sup_{K \in \mathfrak{K}} \int^* |f| \chi_K d\mu = \sup_{K \in \mathfrak{K}} N_1^\mu(f \chi_K), \tag{1.4}$$

wenn hierbei $\mathfrak{K}$ das System aller kompakten Teilmengen von T bezeichnet. Es gilt dann wieder der Satz: Eine reelle Funktion f auf T ist dann und nur dann im wesentlichen integrierbar, wenn sie meßbar und $\overline{N}_1^\mu(f) < +\infty$ ist. In Analogie zu (1.3) gilt für jede Funktion $f \geq 0$ aus $\overline{\mathscr{L}}^1(\mu)$:

$$\int f d\mu = \overline{N}_1^\mu(f). \tag{1.5a}$$

Für eine beliebige Funktion $f \in \overline{\mathscr{L}}^1(\mu)$ gilt:

$$\left|\int f d\mu\right| \leq \overline{N}_1^\mu(f) = \int |f| d\mu. \tag{1.5b}$$

Ferner ist $f \to \int f d\mu$ die einzige, bezüglich der durch $\overline{N}_1^\mu$ definierten Topologie stetige Linearform auf $\overline{\mathscr{L}}^1(\mu)$, welche auf dem Unterraum $\mathscr{K}(T)$ mit dem Maß μ, d. h. mit der Linearform $f \to \mu(f)$ übereinstimmt.

Für auf T *stetige*, reelle Funktionen f fallen übrigens die Begriffe „integrierbar" und „im wesentlichen integrierbar" zusammen. Dies folgt aus N. Bourbaki [6], § 2, prop. 2, wonach für solche Funktionen $\overline{N}_1^\mu(f) = N_1^\mu(f)$ ist.

Jede Menge $A \subset T$ mit $\chi_A \in \overline{\mathscr{L}}^1(\mu)$ wird im wesentlichen (μ-)integrierbar und $\mu(A) = \int \chi_A d\mu$ das Maß von A genannt. Es erweist sich A genau dann als lokal-vernachlässigbar, wenn A im wesentlichen integrierbar und $\mu(A) = 0$ ist.

1.4. *Massentilgung und induziertes Maß.* — Es sei wieder μ ein positives Maß auf einem lokal-kompakten Raum T. Man sagt, es wird μ von einer Menge $M \subset T$ *getragen* oder es ist μ auf M *konzentriert*, wenn die Komplementärmenge $\complement M = T - M$ bezüglich μ lokal-vernachlässigbar ist.

Ist S eine μ-meßbare Teilmenge von T, so ist für jede Funktion $f \in \mathscr{K}(T)$ die Funktion $f \chi_S$ integrierbar und daher $f \to \int f \chi_S d\mu$ eine positive Linearform auf $\mathscr{K}(T)$, also ein positives Maß auf T. Wir bezeichnen dieses mit $\chi_S \mu$ und nennen es das aus μ durch *Tilgung der Masse außerhalb* S entstehende Maß. Gerechtfertigt wird diese Benennung durch die Bemerkung, daß $\chi_S \mu$ auf S konzentriert ist. Ein Maß $\nu \in \mathscr{M}(T)$ wird genau dann von einer Menge $M \subset T$ getragen, wenn M meßbar und $\nu = \chi_M \nu$ ist.

Zur Frage der Integrierbarkeit bezüglich des Maßes $\chi_S \mu$ sei an das folgende Resultat erinnert. Eine reelle Funktion f auf T ist dann und nur dann im wesentlichen $(\chi_S \mu)$-integrierbar, wenn die Funktion $f \chi_S$ im wesentlichen μ-integrierbar ist. Es gilt dann

$$\int f d(\chi_S \mu) = \int f \chi_S d\mu. \tag{1.6}$$

Das Maß $\chi_S \mu$ ist insbesondere definiert für jeden lokal-kompakten Unterraum S von T, da jeder solche als Durchschnitt einer in T abgeschlossenen und einer in T offenen Menge dargestellt werden kann und somit meßbar ist. Dann aber ist $\chi_S \mu$ wohl zu unterscheiden von dem von μ in einem solchen

lokal-kompakten Unterraum *induzierten* Maß μ_S. Während $\chi_S\mu$ ein Maß auf T ist, handelt es sich bei μ_S um ein positives Maß auf S, welches wie folgt definiert ist: Für jede Funktion $g \in \mathscr{K}(S)$ ist diejenige Fortsetzung g' auf T μ-integrierbar, welche auf $T - S$ gleich Null ist. Es ist $\int g d\mu_S = \int g' d\mu$ für jedes $g \in \mathscr{K}(S)$. Wegen $g' = g' \chi_S$ induzieren μ und $\chi_S\mu$ dasselbe Maß μ_S in S. Zu einem beliebigen Maß $\nu \in \mathscr{M}(S)$ gibt es im allgemeinen kein Maß $\mu \in \mathscr{M}(T)$ mit $\nu = \mu_S$. Wir werden hierauf in der nächsten Nummer zurückkommen.

Schließlich führen wir noch folgende Bezeichnung ein: Für eine Teilmenge A von T bezeichnen wir mit $\mathscr{M}(T, A)$ die Menge aller auf A konzentrierten Maße aus $\mathscr{M}(T)$. Demnach ist speziell $\mathscr{M}(T, T) = \mathscr{M}(T)$.

1.5. *Bilder von Maßen*[9]). — Es sei $\varphi: X \to Y$ eine beliebige Abbildung eines lokal-kompakten Raumes X in einen lokal-kompakten Raum Y. Dann ordnet φ gewissen Maßen $\mu \in \mathscr{M}(X)$ ein Maß $\varphi(\mu) \in \mathscr{M}(Y)$ als Bild zu. Und zwar ist das *Bildmaß* $\varphi(\mu)$ für genau diejenigen Maße $\mu \in \mathscr{M}(X)$ definiert, für welche die Funktion $f \circ \varphi$ für jedes $f \in \mathscr{K}(Y)$ im wesentlichen μ-integrierbar ist. $\varphi(\mu)$ wird dann erklärt durch die Gleichung

$$\int f d\,\varphi(\mu) = \int f \circ \varphi \, d\mu \quad \text{für alle } f \in \mathscr{K}(Y)\,. \tag{1.7}$$

Man zeigt, daß $\varphi(\mu)$ dann und nur dann definiert ist, wenn die Abbildung φ *μ-eigentlich*, d. h. wenn das Urbild $\overset{-1}{\varphi}(K)$ einer jeden in Y kompakten Menge K im wesentlichen μ-integrierbar ist.

Die Gleichung (1.7) gilt allgemeiner für jede im wesentlichen $\varphi(\mu)$-integrierbare reelle Funktion f auf Y, sofern $\varphi(\mu)$ überhaupt im Sinne der obigen Definition existiert. Genauer ist eine reelle Funktion f auf Y dann und nur dann im wesentlichen $\varphi(\mu)$-integrierbar, wenn $f \circ \varphi$ im wesentlichen μ-integrierbar ist.

Für eine μ-eigentliche Abbildung $\varphi: X \to Y$ ($\mu \in \mathscr{M}(X)$) und eine beliebige Abbildung $\psi: Y \to Z$ von Y in einen lokal-kompakten Raum Z gilt das folgende *Transitivitätsgesetz*: Es ist ψ dann und nur dann $\varphi(\mu)$-eigentlich, wenn die zusammengesetzte Abbildung $\psi \circ \varphi: X \to Z$ μ-eigentlich ist. Es gilt dann: $(\psi \circ \varphi)(\mu) = \psi(\varphi(\mu))$.

Die gegen Ende der letzten Nummer offengelassene Frage kann jetzt folgendermaßen beantwortet werden: Es sei S ein lokal-kompakter Unterraum eines lokal-kompakten Raumes T und $j: S \to T$ die kanonische Injektion von S in T. Ein Maß $\nu \in \mathscr{M}(S)$ wird dann und nur dann von einem Maß aus $\mathscr{M}(T)$ in S induziert, wenn das Bildmaß $j(\nu)$ definiert, wenn also die Injektion j ν-eigentlich ist. Dann wird nämlich ν durch $j(\nu)$ in S induziert.

Schließlich ist für uns noch von Interesse die Frage nach der algebraischen Struktur (in $\mathscr{M}(X)$) des Definitionsbereiches $\mathscr{E}(\varphi)$ der durch eine Abbildung $\varphi: X \to Y$ vermittelten Abbildung aus $\mathscr{M}(X)$ in $\mathscr{M}(Y)$. Wir beweisen hierzu:

Hilfssatz 1. *Es sei $\varphi: X \to Y$ eine Abbildung eines lokal-kompakten Raumes X in einen lokal-kompakten Raum Y; μ und ν seien positive Maße auf X. Dann gilt:* (a) *mit $\varphi(\mu)$ und $\varphi(\nu)$ ist auch $\varphi(\mu + \nu)$ definiert;* (b) *mit*

[9]) Vgl. hierzu insbesondere N. Bourbaki [6], § 6 und § 7.

$\varphi(\mu)$ ist $\varphi(\alpha\mu)$ für jede Zahl $\alpha \geqq 0$ definiert; (c) *ist $\nu \leqq \mu$ und ist $\varphi(\mu)$ definiert, so ist auch $\varphi(\nu)$ definiert. In den einzelnen Fällen gilt ferner:*

$$(1.8)\quad \begin{array}{ll} \text{(a)} & \varphi(\mu+\nu) = \varphi(\mu) + \varphi(\nu)\,; \\ \text{(b)} & \varphi(\alpha\mu) = \alpha\varphi(\mu)\,; \\ \text{(c)} & \varphi(\nu) \leqq \varphi(\mu)\,. \end{array}$$

Beweis. Zu (a): Die Behauptung ist offenbar bewiesen, wenn gezeigt werden kann, daß $\mathcal{D} = \overline{\mathcal{L}}^1(\mu) \cap \overline{\mathcal{L}}^1(\nu)$ eine Teilmenge von $\overline{\mathcal{L}}^1(\mu+\nu)$ und $\int f d(\mu+\nu) = \int f d\mu + \int f d\nu$ für jedes $f \in \mathcal{D}$ ist. Bekannt ist (vgl. M. Morse u. W. Transue [12], theorem 6.1) die Relation $\mathcal{L}^1(\mu) \cap \mathcal{L}^1(\nu) \subset \mathcal{L}^1(\mu+\nu)$, aus der unmittelbar folgt, daß jede zugleich μ- und ν-meßbare Menge oder reelle Funktion auch $(\mu+\nu)$-meßbar ist. Ferner ist bekannt (vgl. N. Bourbaki [5], p. 113, prop. 15) die Gleichung $N_1^{\mu+\nu} = N_1^\mu + N_1^\nu$, aus der bei Beachtung von (1.4) leicht die entsprechende Gleichung $\overline{N}_1^{\mu+\nu} = \overline{N}_1^\mu + \overline{N}_1^\nu$ folgt. Nun ist nach Nr. 1.3 jede Funktion $f \in \mathcal{D}$ μ- und ν-meßbar, also nach dem ersten Ergebnis $(\mu+\nu)$-meßbar; ferner sind $\overline{N}_1^\mu(f)$ und $\overline{N}_1^\nu(f)$ endlich, woraus nach dem zweiten Ergebnis die Endlichkeit von $\overline{N}_1^{\mu+\nu}(f)$ folgt. Abermals nach Nr. 1.3 ist dann aber f ein Element von $\overline{\mathcal{L}}^1(\mu+\nu)$. Für jede Funktion $f \in \mathcal{D}$ ist nach (1.5b) weiter

$$\left|\int f d\mu + \int f d\nu\right| \leqq \left|\int f d\mu\right| + \left|\int f d\nu\right| \leqq \overline{N}_1^\mu(f) + \overline{N}_1^\nu(f) = \overline{N}_1^{\mu+\nu}(f)\,,$$

also die Linearform $f \to \int f d\mu + \int f d\nu$ auf $\mathcal{D}$ stetig bezüglich der durch $\overline{N}_1^{\mu+\nu}$ definierten Topologie. Auf dem linearen Unterarm $\mathcal{K}(X)$ von $\mathcal{D}$ ist diese Linearform gleich $\mu+\nu$. Hieraus folgt nach dem in Nr. 1.3 Gesagten die behauptete Gleichheit $\int f d(\mu+\nu) = \int f d\mu + \int f d\nu$.

Zu (b): Für $\alpha = 0$ ist die Behauptung trivial. Für $\alpha > 0$ geht man von der bei N. Bourbaki (loc. cit.) bewiesenen Gleichheit $N_1^{\alpha\mu} = \alpha N_1^\mu$ aus und beweist der Reihe nach durch Zurückgehen auf die entsprechenden Definitionen mühelos: $\overline{N}_1^{\alpha\mu} = \alpha\overline{N}_1^\mu$, $\overline{\mathcal{L}}^1(\alpha\mu) = \overline{\mathcal{L}}^1(\mu)$ sowie $\int f d(\alpha\mu) = \alpha\int f d\mu$ für jedes $f \in \overline{\mathcal{L}}^1(\mu)$. Hieraus folgt dann die Behauptung.

Zu (c): Nach N. Bourbaki (loc. cit.) ist $N_1^\nu \leqq N_1^\mu$, woraus $\overline{N}_1^\nu \leqq \overline{N}_1^\mu$ und weiter $\overline{\mathcal{L}}^1(\mu) \subset \overline{\mathcal{L}}^1(\nu)$ folgt. Mit $\varphi(\mu)$ ist daher auch $\varphi(\nu)$ definiert. Weiter ist μ darstellbar in der Form $\mu = \nu + \varrho$ mit $\varrho \in \mathcal{M}(X)$; wegen $\varrho \leqq \mu$ ist nach dem soeben Bewiesenen auch $\varphi(\varrho)$ definiert. Aus der Behauptung (a) folgt daher $\varphi(\mu) = \varphi(\nu) + \varphi(\varrho)$ und hieraus $\varphi(\nu) \leqq \varphi(\varrho)$.

Aus diesem Hilfssatz ergibt sich nun sofort die Antwort auf unsere Frage:

Korollar. *Die Menge $\mathcal{E} = \mathcal{E}(\varphi)$ aller Maße $\mu \in \mathcal{M}(X)$, für welche das Bildmaß $\varphi(\mu)$ definiert ist, ist ein konvexer, erblicher Unterkegel von $\mathcal{M}(X)$. D. h. $\mathcal{E}$ besitzt folgende Eigenschaften:*

$$(1.9)\quad \begin{array}{c} \mu, \nu \in \mathcal{E} \Rightarrow \mu + \nu \in \mathcal{E}\,; \\ \mu \in \mathcal{E} \Rightarrow \alpha\mu \in \mathcal{E} \quad (\alpha \geqq 0, \alpha \in \mathbf{R})\,; \\ \mu \in \mathcal{E}, \nu \leqq \mu \Rightarrow \nu \in \mathcal{E} \ (\nu \in \mathcal{M}(X))\,. \end{array}$$

1.6. *Konservative Abbildungen.* — Es sei wieder $\varphi: X \to Y$ eine Abbildung eines lokal-kompakten Raumes X in einen lokal-kompakten Raum Y. Für

die charakteristische Funktion $f = \chi_A$ einer Menge $A \subset Y$ folgt aus (1.7) und dem im Anschluß an diese Formel erwähnten Satz: Die Menge A ist genau dann lokal-vernachlässigbar bezüglich des Bildmaßes $\varphi(\mu)$ eines Maßes $\mu \in \mathscr{E}(\varphi)$, wenn die Menge $\overset{-1}{\varphi}(A)$ lokal-vernachlässigbar bezüglich μ ist. Speziell für die Menge $A = Y - \varphi(X)$, deren Urbild leer ist, gilt also: $Y - \varphi(X)$ ist $\varphi(\mu)$-lokal-vernachlässigbar, d. h. es wird $\varphi(\mu)$ für jedes Maß $\mu \in \mathscr{E}(\varphi)$ von $\varphi(X)$ getragen. Somit definiert φ genauer eine Abbildung von $\mathscr{E}(\varphi)$ in $\mathscr{M}(Y, \varphi(X))$.

Hauptanliegen der vorliegenden Untersuchung ist die Frage nach Abbildungen φ, welche eine Abbildung von $\mathscr{E}(\varphi)$ auf $\mathscr{M}(Y, \varphi(X))$ vermitteln. Wir definieren daher:

Definition 1. *Es sei $\varphi : X \to Y$ eine Abbildung eines lokal-kompakten Raumes X in einen lokal-kompakten Raum Y. Ein Maß $\nu \in \mathscr{M}(Y)$ heiße **darstellbar** bezüglich φ (kürzer: (φ-)darstellbar), wenn mindestens ein Maß $\mu \in \mathscr{M}(X)$ existiert, für welches $\varphi(\mu)$ definiert und $\varphi(\mu) = \nu$ ist.*

*Die Abbildung φ heiße **konservativ**, wenn jedes auf $\varphi(X)$ konzentrierte, positive Maß auf Y φ-darstellbar ist.*

Beispiel. Es sei X ein lokal-kompakter Unterraum eines lokal-kompakten Raumes Y und $j : X \to Y$ die kanonische *Injektion* von X in Y. Dann ist die Abbildung j konservativ, denn für jedes Maß $\nu \in \mathscr{M}(Y, X) = \mathscr{M}(Y, j(X))$ ist das von ν in X induzierte Maß ν_X ein Maß aus $\mathscr{M}(X)$ mit $j(\nu_X) = \nu$. (Vgl. N. Bourbaki [6], § 7, prop. 3.)

In den folgenden Paragraphen werden wir wichtige Klassen *stetiger*, konservativer Abbildungen kennenlernen. Zunächst soll jedoch die Existenz stetiger, nicht-konservativer Abbildungen durch Angabe eines Beispiels nachgewiesen werden.

1.7. *Beispiel einer stetigen, nicht-konservativen Abbildung.* — Es sei $I = [0,1]$ die Menge aller reellen Zahlen x mit $0 \leqq x \leqq 1$. Der Raum X bzw. Y entsteht, indem man der Menge I die diskrete bzw. die übliche (euklidische) Topologie aufprägt. Beide Räume sind lokal-kompakt (Y ist sogar kompakt); die identische Abbildung φ von I auf sich ist eine stetige Abbildung von X *auf* Y. Wir behaupten:

Ein Maß $\nu \in \mathscr{M}(Y)$ ist dann und nur dann φ-darstellbar, wenn es die Gestalt $\nu = \sum_{n=1}^{\infty} a_n \varepsilon_{x_n}$[4]) besitzt, wobei $(a_n)_{n=1,2,\ldots}$ eine Folge reeller Zahlen $\geqq 0$ mit $\sum_{n=1}^{\infty} a_n < +\infty$ und $(x_n)_{n=1,2,\ldots}$ eine Folge von Punkten aus I ist.

Beweis. Es sei $\nu \in \mathscr{M}(Y)$ ein φ-darstellbares Maß und μ ein Maß aus $\mathscr{M}(X)$ mit $\varphi(\mu) = \nu$. Da auf einem diskreten Raum jedes Maß diskret ist (N. Bourbaki [5], p. 51), gibt es eine reelle Funktion $\alpha \geqq 0$ auf X derart, daß $\int g\,d\mu = \sum_{x \in I} \alpha(x)\, g(x)$ ist für jedes $g \in \mathscr{K}(X)$. Eine reelle Funktion h auf X ist

[4]) Die Konvergenz dieser Reihe ist im Sinne der vagen Topologie zu verstehen.

genau dann μ-integrierbar, wenn die Familie $(\alpha(x)h(x))_{x\in I}$ in $\mathbf{R}$ absolut summierbar oder, was hiermit gleichwertig ist[6]), summierbar ist; das Integral errechnet sich dann zu $\int h d\mu = \sum_{x\in I} \alpha(x)h(x)$ (nach [5], p.146). Wegen $\varphi(\mu) = \nu$ und der Stetigkeit von φ, die das Übereinstimmen der Begriffe „integrierbar" und „im wesentlichen integrierbar" für alle Funktionen $f \circ \varphi$ mit $f \in \mathcal{K}(Y)$ zur Folge hat, ist also für jedes $f \in \mathcal{K}(Y)$ die Familie $(\alpha(x)f(x))_{x\in I}$ summierbar und $\int f d\nu = \sum_{x\in I} \alpha(x)f(x)$. Da die konstante Funktion $f = 1$ zu $\mathcal{K}(Y)$ gehört, folgt hieraus die Summierbarkeit von $(\alpha(x))_{x\in I}$. D. h. aber: es gibt eine Folge $(x_n)_{n=1,2,\ldots}$ von Punkten aus I derart, daß $\alpha(x) = 0$ ist für alle $x \neq x_n$, $x \in I$, und daß die Reihe $\sum \alpha(x_n)$ konvergiert. Für jedes $f \in \mathcal{K}(Y)$ ist dann $\int f d\nu = \sum \alpha(x_n) f(x_n)$ und somit $\nu = \sum \alpha(x_n)\,\varepsilon_{x_n}$. Somit hat ν die behauptete Gestalt. — Umgekehrt ist jedes Maß $\nu = \sum a_n \varepsilon_{x_n}$ aus $\mathcal{M}(Y)$ mit $a_n \geq 0$ und $\sum a_n < +\infty$ darstellbar. Es ist nämlich $\mu = \sum a_n \varepsilon_{x_n}$, wobei jetzt ε_{x_n} als Maß auf X zu interpretieren ist, ein Maß aus $\mathcal{M}(X)$ mit $\varphi(\mu) = \nu$.

Aus dem soeben Bewiesenen folgt nun sofort, daß die Abbildung φ nicht konservativ sein kann. Ist nämlich $\nu = \sum a_n \varepsilon_{x_n}$ ein darstellbares Maß aus $\mathcal{M}(Y)$, so ist (abermals nach [5], p. 146) $\nu(\{x\}) = 0$ für alle $x \neq x_n$, $x \in I$, und $\nu(\{x_n\}) = a_n$ für jedes n. Ein diffuses Maß $\nu \neq 0$ aus $\mathcal{M}(Y)$, also beispielsweise das Lebesguesche Maß auf Y, kann somit nicht darstellbar sein.

§ 2. Ein Lokalisationsprinzip

Von nun an wird unser Interesse den *stetigen*, konservativen Abbildungen gelten. Es soll hier zunächst gezeigt werden, daß die Frage nach der Darstellbarkeit eines Maßes $\nu \in \mathcal{M}(Y)$ bezüglich einer stetigen Abbildung $\varphi: X \to Y$ ein lokales Problem ist. Zu diesem Zweck führen wir folgenden Begriff ein:

Definition. *Es sei $\varphi: X \to Y$ eine stetige Abbildung eines lokal-kompakten Raumes X in einen lokal-kompakten Raum Y. Wir nennen ein Maß $\nu \in \mathcal{M}(Y)$ (φ-)lokal-darstellbar, wenn jeder Punkt $y \in \varphi(X)$ eine offene Umgebung V in Y besitzt mit folgender Eigenschaft: Das durch ν in V induzierte Maß $\nu_V \in \mathcal{M}(V)$ ist φ'_V-darstellbar, wenn hierbei $\varphi'_V: U \to V$ diejenige Abbildung bezeichnet, welche auf $U = \overset{-1}{\varphi}(V)$ mit φ übereinstimmt.*

Bezeichnen wir mit $j_U: U \to X$ bzw. $j_V: V \to Y$ die kanonische Injektion von U in X bzw. von V in Y, so ist also φ'_V die einzige Abbildung von U in V, für welche das folgende Diagramm kommutativ ist:

$$(2.1)\qquad \begin{array}{ccc} U & \xrightarrow{\varphi'_V} & V \\ \downarrow{\scriptstyle j_U} & & \downarrow{\scriptstyle j_V} \\ X & \xrightarrow[\varphi]{} & Y \end{array}$$

Zu der Definition ist weiter zu bemerken, daß U und V als offene Mengen in lokal-kompakten Räumen selbst wieder lokal-kompakte Räume sind. Ferner

[6]) Vgl. hierzu N. BOURBAKI [8], chap. III, § 7, théorème 3.

ist die Abbildung φ'_V zu unterscheiden von der Restriktion $\varphi_U : U \to Y$ von φ auf U. Es wäre sinnlos, von der φ_U-Darstellbarkeit von ν_V zu sprechen, da ν_V kein Maß auf Y ist. Es gilt jedoch folgender

Hilfssatz 2. *Es sei $\varphi : X \to Y$ eine stetige Abbildung eines lokal-kompakten Raumes X in einen lokal-kompakten Raum Y, T ein lokal-kompakter Unterraum von Y, $S = \overset{-1}{\varphi}(T)$ und ν ein positives Maß auf Y. Dann sind folgende drei Aussagen gleichwertig:*

(a) *Das durch ν in T induzierte Maß $\nu_T \in \mathscr{M}(T)$ ist darstellbar bezüglich der Abbildung $\varphi'_T : S \to T$, die auf S mit φ übereinstimmt.*

(b) *Das aus ν durch Tilgung der Masse außerhalb T entstehende Maß $\tau = \chi_T \nu \in \mathscr{M}(Y)$ ist φ-darstellbar.*

(c) *Das Maß τ ist darstellbar bezüglich der Restriktion $\varphi_S : S \to Y$ von φ auf S.*

Beweis. Die Aussagen (a) und (c) sind rein formal betrachtet sinnvoll, da S ein lokal-kompakter Unterraum von X ist. Dies folgt aus dem bekannten Satz, wonach ein Unterraum eines lokal-kompakten Raumes genau dann lokal-kompakt ist, wenn er Durchschnitt einer abgeschlossenen und einer offenen Menge ist. Wir bezeichnen mit $j_S : S \to X$ bzw. $j_T : T \to Y$ die kanonische Injektion von S in X bzw. T in Y.

(a) $\Rightarrow$ (b). Es sei $\sigma' \in \mathscr{M}(S)$ ein Maß mit $\varphi'_T(\sigma') = \nu_T$. Wir zeigen zunächst, daß die Abbildung j_S dann σ'-eigentlich, also das Bildmaß $j_S(\sigma')$ definiert ist. Es sei hierzu K eine in X kompakte Menge. Wegen der Stetigkeit von φ ist dann $L = \varphi(K)$ kompakt in Y; da ν_T durch ν in T induziert wird, ist folglich die Spur $L \cap T$ von L in T im wesentlichen ν_T-integrierbar (vgl. Nr. 1.5). Ebenfalls nach Nr. 1.5 ist dann wegen $\varphi'_T(\sigma') = \nu_T$ auch die Menge $(\varphi'_T)^{-1}(L \cap T) = \overset{-1}{\varphi}(L) \cap S$ im wesentlichen σ'-integrierbar. Nun ist $K \cap S \subset \overset{-1}{\varphi}(L) \cap S$ und in S abgeschlossen, also σ'-meßbar. Da $\overset{-1}{\varphi}(L) \cap S$ im wesentlichen σ'-integrierbar ist, folgt dann endlich nach Nr. 1.3, daß auch $K \cap S$ im wesentlichen σ'-integrierbar ist, und zwar für jede in X kompakte Menge K. Somit ist $\sigma = j_S(\sigma')$ definiert und ein Element von $\mathscr{M}(X)$. Nach dem Beispiel der Nr. 1.6 ist weiter $j_T(\nu_T) = \tau$. Aus $\varphi'_T(\sigma') = \nu_T$ und $j_T(\nu_T) = \tau$ folgt aber nach dem Transitivitätsgesetz für Bildmaße (Nr. 1.6) die Gleichung $(j_T \circ \varphi'_T)(\sigma') = \tau$. Zufolge der Kommutativität des Diagramms (2.1) (mit S bzw. T anstelle von U bzw. V) ist $j_T \circ \varphi'_T = \varphi \circ j_S$ und somit $(\varphi \circ j_S)(\sigma') = \tau$. Weil wir festgestellt haben, daß $\sigma = j_S(\sigma')$ definiert ist, liefert eine erneute Anwendung des Transitivitätsgesetzes für Bildmaße die Aussage, daß $\varphi(\sigma)$ definiert und $\varphi(\sigma) = \tau$ ist. Somit ist das Maß τ φ-darstellbar.

(b) $\Rightarrow$ (c). Es sei $\sigma \in \mathscr{M}(X)$ ein Maß mit $\varphi(\sigma) = \tau$. Die Menge $Y - T$ ist τ-lokal-vernachlässigbar; aus (1.7) (mit $\mu = \sigma$ und $f = \chi_{Y-T}$) folgt daher, daß $\overset{-1}{\varphi}(Y - T) = X - S$ σ-lokal-vernachlässigbar ist, also σ von S getragen wird. Nach Nr. 1.6 gilt daher $j_S(\sigma_S) = \sigma$ für das von σ in S induzierte Maß σ_S. Da außerdem $\varphi(\sigma) = \tau$ ist, liefert das Transitivitätsgesetz für Bildmaße: $(\varphi \circ j_S)(\sigma_S) = \tau$. Es ist aber $\varphi_S = \varphi \circ j_S$ und somit das Maß τ φ_S-darstellbar.

(c) $\Rightarrow$ (a). Es sei $\sigma' \in \mathscr{M}(S)$ ein Maß mit $\varphi_S(\sigma') = \tau$. Dann folgt hieraus $\varphi'_T(\sigma') = \nu_T$ und damit (a). In der Tat: für jede Funktion $f \in \mathscr{K}(T)$ und deren

Fortsetzung t' auf Y, die auf $Y - T$ gleich Null ist, gilt nach Definition des induzierten Maßes ν_T sowie (1.6):

$$\int t\,d\nu_T = \int t'\,d\nu = \int t'\,\chi_T d\nu = \int t'\,d\tau\,.$$

Wegen $\varphi_S(\sigma') = \tau$ ist nach (1.7) weiter:

$$\int t'\,d\tau = \int t' \circ \varphi_S\,d\sigma'\,.$$

Für jedes $t \in \mathcal{K}(T)$ ist daher die auf S definierte Funktion $t \circ \varphi'_T = t' \circ \varphi_S$ im wesentlichen σ'-integrierbar (wegen der Stetigkeit von $t \circ \varphi'_T$ übrigens sogar σ'-integrierbar, vgl. Nr. 1.3) und es gilt $\int t\,d\nu_T = \int t \circ \varphi'_T d\sigma'$. Somit ist $\varphi'_T(\sigma') = \nu_T$ und der Hilfssatz damit bewiesen.

Nunmehr sind wir in der Lage, das zu Beginn dieses Paragraphen angekündigte Lokalisationsprinzip zu formulieren und zu beweisen.

Satz 1 *(Lokalisationsprinzip). Es sei $\varphi: X \to Y$ eine stetige Abbildung eines lokal-kompakten Raumes X in einen lokal-kompakten Raum Y. Ein positives Maß ν auf Y ist dann und nur dann φ-darstellbar, wenn ν auf $\varphi(X)$ konzentriert und φ-lokal-darstellbar ist.*

Beweis. Sei $\nu \in \mathcal{M}(Y)$ ein darstellbares Maß. Wie in Nr. 1.6 bemerkt wurde, ist dann ν auf $\varphi(X)$ konzentriert. Das Maß ist lokal-darstellbar, denn offenbar leistet für jeden Punkt $y \in \varphi(X)$ die offene Umgebung $V = Y$ das in Definition 2 Verlangte. Zu beweisen ist daher nur die Aussage „dann".

Also sei jetzt ν ein lokal-darstellbares Maß aus $\mathcal{M}(Y, \varphi(X))$. Für jede in Y lokal-kompakte Menge Z bezeichne $\nu'_Z = \chi_Z \nu$ das aus ν durch Tilgung der Masse außerhalb Z entstehende Maß aus $\mathcal{M}(Y)$. Mit $\mathfrak{S}$ werde die Menge aller Paare (G, μ) bezeichnet, in welchen G eine offene Teilmenge von Y und μ ein Maß aus $\mathcal{M}(X)$ mit $\varphi(\mu) = \nu'_G$ ist. $\mathfrak{S}$ ist dann nicht leer, denn es ist z. B. $(\emptyset, 0) \in \mathfrak{S}$. In $\mathfrak{S}$ definieren wir wie folgt eine Relation $\prec$: Die Relation $(G_1, \mu_1) \prec (G_2, \mu_2)$ bestehe zwischen zwei Elementen von $\mathfrak{S}$ genau dann, wenn $G_1 \subset G_2$ und $\mu_1 \leq \mu_2$ ist. Offenbar ist $\prec$ eine Ordnungsrelation in $\mathfrak{S}$.

Wir wollen zeigen, daß $\mathfrak{S}$ durch $\prec$ sogar induktiv geordnet ist. Hierzu sei $\mathfrak{F}$ eine totalgeordnete, nicht-leere Teilmenge von $\mathfrak{S}$. Es bezeichne $\mathfrak{F}_Y$ das System aller Mengen $G \subset Y$, zu denen ein $\mu \in \mathcal{M}(X)$ mit $(G, \mu) \in \mathfrak{F}$ existiert; $\mathfrak{F}_{\mathcal{M}}$ sei das System aller Maße $\mu \in \mathcal{M}(X)$, zu denen eine Menge G mit $(G, \mu) \in \mathfrak{F}$ existiert. Dann ist die Menge

$$H = \bigcup_{G \in \mathfrak{F}_Y} G \tag{2.2}$$

offen in Y und $\mathfrak{F}_{\mathcal{M}}$ eine totalgeordnete Teilmenge von $\mathcal{M}(X)$. Weiter ist $\mathfrak{F}_{\mathcal{M}}$ beschränkt, d. h.

$$\sup_{\mu \in \mathfrak{F}_{\mathcal{M}}} \left|\int g\,d\mu\right| < +\infty \quad \text{für jedes } g \in \mathcal{K}(X)\,.$$

Zu jedem $g \in \mathcal{K}(X)$ gibt es nämlich ein $f \in \mathcal{K}_+(Y)$ mit $|g| \leq f \circ \varphi$. Man hat dazu nur zu bemerken, daß das Bild $\varphi(T_g)$ des Trägers T_g von g kompakt ist und somit eine Funktion $f' \in \mathcal{K}_+(Y)$ existiert mit $f'(y) = 1$ für alle $y \in \varphi(T_g)$. Dann leistet $f = \alpha f'$ das Verlangte für hinreichend großes $\alpha > 0$. Ist aber

$f \in \mathcal{K}_+(Y)$ derart gewählt, so folgt die behauptete Beschränktheit von $\mathfrak{F}_{\mathcal{M}}$ aus der Abschätzung:

$$|\int g d\mu| \leq \int |g|\, d\mu \leq \int f \circ \varphi\, d\mu = \int f d\nu'_G \leq \int f d\nu \quad (\mu \in \mathfrak{F}_{\mathcal{M}}).$$

Da $\mathfrak{F}_{\mathcal{M}}$ eine beschränkte und totalgeordnete Teilmenge von $\mathcal{M}(X)$ ist, konvergiert die Filterbasis der Abschnitte von $\mathfrak{F}_{\mathcal{M}}$ gegen ein Maß $\eta \in \mathcal{M}(X)$ im Sinne der vagen Topologie. Nach dem Monotonieprinzip existiert nämlich $\lim_{\mu \in \mathfrak{F}_{\mathcal{M}}} \int g d\mu$ für jede Funktion $g \in \mathcal{K}_+(X)$ und damit für jedes $g \in \mathcal{K}(X)$. Folglich ist

(2.3) $$\int g d\eta = \sup_{\mu \in \mathfrak{F}_{\mathcal{M}}} \int g d\mu = \lim_{\mu \in \mathfrak{F}_{\mathcal{M}}} \int g d\mu \quad \text{für jedes } g \in \mathcal{K}_+(X).$$

Wir behaupten, daß das Paar (H, η) in $\mathfrak{S}$ liegt. Um zu zeigen, daß $\varphi(\eta)$ definiert ist, genügt es zu beweisen, daß die Funktion $f \circ \varphi$ für jedes $f \in \mathcal{K}_+(Y)$ η-integrierbar ist. Es sei hierzu $g \in \mathcal{K}_+(X)$ mit $0 \leq g \leq f \circ \varphi$; dann gilt

$$\int g d\mu \leq \int f \circ \varphi\, d\mu = \int f d\nu'_G \leq \int f d\nu'_H \quad \text{für jedes } (G, \mu) \in \mathfrak{F}$$

und folglich nach (2.3)

$$\int g d\eta = \sup_{\mu \in \mathfrak{F}_{\mathcal{M}}} \int g d\mu \leq \int f d\nu'_H .$$

Bei Beachtung von (1.1a) und (1.2) ergibt dies:

$$N_1^\eta(f \circ \varphi) \leq \int f d\nu'_H \qquad (f \in \mathcal{K}_+(Y)).$$

Da $f \circ \varphi$ als stetige Funktion meßbar und nach der vorausgehenden Ungleichung $N_1^\eta(f \circ \varphi)$ endlich ist, so ist $f \circ \varphi$ η-integrierbar und

(2.4) $$\int f \circ \varphi d\eta \leq \int f d\nu'_H \qquad (f \in \mathcal{K}_+(Y)).$$

Es ist $\varphi(\eta) = \nu'_H$ und damit $(H, \eta) \in \mathfrak{S}$, wenn gezeigt werden kann, daß in (2.4) das Gleichheitszeichen steht. Aus (2.3) folgt zunächst $\mu \leq \eta$ für jedes $\mu \in \mathfrak{F}_{\mathcal{M}}$ und hieraus

$$\int f \chi_G d\nu = \int f d\nu'_G = \int f \circ \varphi d\mu \leq \int f \circ \varphi d\eta$$

für jedes Paar $(G, \mu) \in \mathfrak{F}$. Also ist

(2.5) $$\int f \chi_G d\nu \leq \int f \circ \varphi d\eta \qquad \text{für jedes } G \in \mathfrak{F}_Y \text{ und } f \in \mathcal{K}_+(Y).$$

Nun ist $(f \chi_G)_{G \in \mathfrak{F}_Y}$ eine bezüglich $\leq$ totalgeordnete Familie von ν-integrierbaren, nach unten halbstetigen, nicht-negativen Funktionen mit der oberen Einhüllenden $f \chi_H$; nach N. Bourbaki [5], p. 151, gilt daher

$$\int f \chi_H d\nu = \sup_{G \in \mathfrak{F}_Y} \int f \chi_G d\nu \qquad (f \in \mathcal{K}_+(Y)).$$

Zusammen mit (2.5) ergibt dies

(2.6) $$\int f d\nu'_H \leq \int f \circ \varphi d\eta \qquad (f \in \mathcal{K}_+(Y)).$$

Aus (2.4) und (2.6) folgt aber, daß in (2.4) das Gleichheitszeichen steht, also $(H, \eta) \in \mathfrak{S}$ ist. Aus (2.2) und (2.3) folgt schließlich, daß (H, η) eine obere Schranke von $\mathfrak{F}$ in $\mathfrak{S}$ ist. Somit ist $\mathfrak{S}$ induktiv geordnet.

Das Zornsche Lemma sichert nunmehr die Existenz mindestens eines maximalen Elementes (G_0, μ_0) in $\mathfrak{S}$. Wir behaupten, daß $\varphi(X) \subset G_0$ ist. Angenommen, dies ist nicht der Fall. Dann gibt es einen Punkt $y \in \varphi(X) \cap (Y - G_0)$ und zu diesem nach Voraussetzung eine offene Umgebung V derart, daß das

induzierte Maß ν_V darstellbar ist bezüglich der Abbildung $\varphi'_V : U \to V$, die auf $U = \overset{-1}{\varphi}(V)$ mit φ übereinstimmt. Nach dem Hilfssatz 2 ist dann das Maß ν'_V φ-darstellbar, d. h. es existiert ein Maß $\mu \in \mathcal{M}(X)$ mit $(V, \mu) \in \mathfrak{S}$. Wir setzen $T = V \cap (Y - G_0)$; T ist somit ein zu G_0 fremder, lokal-kompakter Unterraum von Y. Wegen der Stetigkeit von φ ist auch $S = \overset{-1}{\varphi}(T)$ lokal-kompakt in X. Folglich sind die aus μ bzw. ν durch Tilgung der Masse außerhalb S bzw. T entstehenden Maße μ'_S bzw. ν'_T definiert. Wir wollen zeigen, daß $\varphi(\mu'_S) = \nu'_T$ ist: Für jede Funktion $f \in \mathcal{K}(Y)$ ist $f\chi_T$ ν'_V-integrierbar, wegen $\varphi(\mu) = \nu'_V$ und (1.7) also $(f\chi_T)\circ\varphi$ im wesentlichen μ-integrierbar und

$$\int (f\chi_T)\circ\varphi\, d\mu = \int f\chi_T d\nu'_V .$$

Da $\nu'_T = \chi_T \nu = \chi_T \chi_V \nu = \chi_T \nu'_V$ und $(f\chi_T)\circ\varphi = (f\circ\varphi)\chi_S$ ist, folgt bei Beachtung von (1.6):

$$\int f\circ\varphi\, d\mu'_S = \int f d\nu'_T \qquad (f \in \mathcal{K}(Y)), \tag{2.7}$$

d. h. es ist $\varphi(\mu'_S) = \nu'_T$. Nun ist aber $H_0 = G_0 \cup V = G_0 \cup T$ offen in Y und $\eta_0 = \mu_0 + \mu'_S$ ein Maß aus $\mathcal{M}(X)$; nach dem Hilfssatz 1 ist $\varphi(\eta_0)$ definiert und

$$\varphi(\eta_0) = \nu'_{G_0} + \nu'_T = (\chi_{G_0} + \chi_T)\nu = \chi_{H_0}\nu = \nu'_{H_0} .$$

Also ist auch (H_0, η_0) ein Element von $\mathfrak{S}$; offenbar ist $G_0 \subset H_0$ und $\mu_0 \leqq \eta_0$, also $(G_0, \mu_0) \prec (H_0, \eta_0)$. Es ist aber $y \in V \cap (Y - G_0)$ und somit $G_0 \neq H_0$, also auch $(G_0, \mu_0) \neq (H_0, \eta_0)$. Dies widerspricht der Maximaleigenschaft von (G_0, μ_0). Also muß unsere Annahme falsch sein: es ist vielmehr $\varphi(X) \subset G_0$.

Schließlich ist nach Voraussetzung ν auf $\varphi(X)$ und damit auch auf G_0 konzentriert, also $\nu'_{G_0} = \nu$. Aus $\varphi(\mu_0) = \nu'_{G_0}$ folgt daher $\varphi(\mu_0) = \nu$, d. h. es ist ν φ-darstellbar. Das war aber zu beweisen.

§ 3. Eigentliche und lokal-eigentliche Abbildungen

3.1. *Konservativität lokal-eigentlicher Abbildungen.* — Von einer wichtigen Klasse stetiger Abbildungen soll nun gezeigt werden, daß sie aus lauter konservativen Abbildungen besteht. Wir erinnern zunächst an die Definition dieser Abbildungen:

Definition 3. *Es sei $\varphi : X \to Y$ eine stetige Abbildung eines lokal-kompakten Raumes X in einen lokal-kompakten Raum Y. Dann heißt φ **eigentlich**, wenn das Urbild $\overset{-1}{\varphi}(K)$ jeder in Y kompakten Menge K kompakt in X ist. Die Abbildung φ heißt **lokal-eigentlich**, wenn jeder Punkt $y \in \varphi(X)$ eine kompakte Umgebung W in Y besitzt, deren Urbild $\overset{-1}{\varphi}(W)$ kompakt in X ist.*

Der Begriff der eigentlichen Abbildung findet sich bei N. Bourbaki [7], der der lokal-eigentlichen Abbildung bei R. Remmert [17].

Jede eigentliche Abbildung ist offensichtlich lokal-eigentlich. Dagegen gilt: *Eine lokal-eigentliche stetige Abbildung $\varphi : X \to Y$ ist dann und nur dann eigentlich, wenn $\varphi(X)$ in Y abgeschlossen ist.* Wenn nämlich $\varphi(X)$ in Y abgeschlossen und φ lokal-eigentlich ist, so zeigt man mit Hilfe des Borelschen Überdeckungssatzes leicht, daß φ eigentlich ist. Die Umkehrung folgt aus einem bekannten Satz (N. Bourbaki [7], p. 99, prop. 13), wonach jede eigentliche Abbildung abgeschlossen ist, also abgeschlossene Mengen in abgeschlossene

Mengen überführt. Für eine stetige Abbildung φ von X *auf* Y sind somit die Begriffe „eigentlich" und „lokal-eigentlich" gleichbedeutend. Eine stetige Abbildung eines kompakten Raumes in einen lokal-kompakten Raum ist stets eigentlich.

Über diese beiden Abbildungstypen beweisen wir zunächst folgenden

Hilfssatz 3. *Für jede stetige Abbildung $\varphi : X \to Y$ eines lokal-kompakten Raumes X in einen lokal-kompakten Raum Y gilt:*

(a) *Die Abbildung φ ist dann und nur dann eigentlich, wenn aus $f \in \mathscr{K}(Y)$ stets $f \circ \varphi \in \mathscr{K}(X)$ folgt.*

(b) *Die Abbildung φ ist dann und nur dann lokal-eigentlich, wenn jeder Punkt $y \in \varphi(X)$ eine offene Umgebung V in Y besitzt, für welche die Abbildung $\varphi'_V : U \to V$ eigentlich ist, welche auf $U = \overset{-1}{\varphi}(V)$ mit φ übereinstimmt.*

Beweis. Zu (a): Trivial ist, daß für eine eigentliche Abbildung die angegebene Bedingung erfüllt ist. Ist umgekehrt diese Bedingung erfüllt, so ist φ eigentlich, da zu jeder kompakten Menge $K \subset Y$ eine Funktion $h \in \mathscr{K}(Y)$ existiert mit $0 \leq h \leq 1$ und $h(y) = 1$ für alle $y \in K$.

Zu (b): Es sei φ lokal-eigentlich. Dann gibt es zu jedem $y \in \varphi(X)$ eine kompakte Umgebung W von y mit kompaktem Urbild $\overset{-1}{\varphi}(W)$. Der offene Kern V von W ist dann eine offene Umgebung von y in Y, die das Verlangte leistet: Für jede in V kompakte Menge K ist aus Stetigkeitsgründen $(\varphi'_V)^{-1}(K) = \overset{-1}{\varphi}(K)$ abgeschlossen, also wegen $\overset{-1}{\varphi}(K) \subset \overset{-1}{\varphi}(W)$ kompakt in X und damit auch in U. Ist umgekehrt V eine offene Umgebung von $y \in \varphi(X)$, für welche die Abbildung φ'_V eigentlich ist, so ist für jede kompakte Umgebung W von y mit $W \subset V$ deren Urbild $\overset{-1}{\varphi}(W) = (\varphi'_V)^{-1}(W)$ kompakt in U und damit in X. Zufolge der lokalen Kompaktheit von X existiert aber mindestens ein solches W.[5a]

[5a]) **Zusatz bei der Korrektur.** Aus Hilfssatz 3 ergibt sich die folgende Kennzeichnung der lokal-eigentlichen Abbildungen: *Eine stetige Abbildung $\varphi : X \to Y$ eines lokal-kompakten Raumes X in einen lokal-kompakten Raum Y ist dann und nur dann lokal-eigentlich, wenn sie eine Faktorisierung $\varphi = j \circ \varphi^*$ besitzt, in welcher $\varphi^* : X \to Y^*$ eine eigentliche Abbildung von X auf einen lokal-kompakten Unterraum Y^* von Y und $j : Y^* \to Y$ die kanonische Injektion von Y^* in Y ist.*

Beweis. „*Nur dann*": Ist $\varphi : X \to Y$ lokal-eigentlich, so ist offenbar das Urbild jeder kompakten Teilmenge von $Y^* = \varphi(X)$ kompakt in X. Zu zeigen ist daher nur, daß (der separierte Raum) Y^* lokal-kompakt ist. Nach Hilfssatz 3 existiert zu einem beliebigen Punkt $y \in Y^*$ eine in Y offene Umgebung V, für welche die zugehörige Abbildung $\varphi'_V : U = \overset{-1}{\varphi}(V) \to V$ eigentlich ist. Da eigentliche Abbildungen abgeschlossen sind, so ist $V' = V \cap Y^* = \varphi'_V(U)$ abgeschlossen in V, also lokal-kompakt. Folglich besitzt y in V' eine kompakte Umgebung W. Es ist aber W auch Umgebung von y in Y^*. In der Tat: es gibt eine in Y offene Menge G mit $y \in G$ und $G \cap V' \subset W$; die Menge $G \cap V$ ist offen in Y und es gilt $G \cap V \cap Y^* = G \cap V'$. Also besitzt jeder Punkt $y \in Y^*$ eine kompakte Umgebung in Y^*, d. h. Y^* ist lokal-kompakt. — „*Dann*": Es sei $\varphi = j \circ \varphi^*$ eine Faktorisierung mit den genannten Eigenschaften. Da $Y^* = \varphi^*(X) = \varphi(X)$ lokal-kompakt ist, gibt es in Y eine abgeschlossene Menge F und eine offene Menge G mit $Y^* = F \cap G$. Jeder Punkt $y \in Y^*$ besitzt in Y eine kompakte Umgebung V mit $V \subset G$. Für diese ist $V \cap Y^* = V \cap F$ kompakt in Y^*, also $\overset{-1}{\varphi}(V) = \varphi^{*-1}(V \cap Y^*)$ kompakt in X. Somit ist φ lokal-eigentlich. — Aus dieser Kennzeichnung folgt übrigens, daß sich im Beweis des folgenden Satzes 2 das Lokalisationsprinzip (Satz 1) noch vermeiden ließe.

Ein Analogon zu der Aussage (a) des Hilfssatzes 3 haben wir in Nr. 1.5 kennengelernt, wonach für eine Abbildung $\varphi: X \to Y$ eines lokalkompakten Raumes in einen ebensolchen Raum das Bild $\varphi(\mu)$ eines Maßes $\mu \in \mathscr{M}(X)$ genau dann definiert ist, wenn die Abbildung μ-eigentlich ist. Hieraus oder aus Hilfssatz 3 folgt übrigens, daß für eine eigentliche (stetige) Abbildung φ das Bild $\varphi(\mu)$ für *jedes* $\mu \in \mathscr{M}(X)$ definiert ist. Im Sinne der Bezeichnungen von 1.5 ist also $\mathscr{E}(\varphi) = \mathscr{M}(X)$ und es definiert φ eine Abbildung von $\mathscr{M}(X)$ in $\mathscr{M}(Y, \varphi(X))$. Daß sogar eine Abbildung auf $\mathscr{M}(Y, \varphi(X))$ vorliegt, lehrt der folgende Satz, aus dem die Bedeutung der eigentlichen und allgemeiner der lokal-eigentlichen Abbildungen für unser Problem erhellt.

Satz 2. *Jede lokal-eigentliche Abbildung $\varphi: X \to Y$ eines lokal-kompakten Raumes X in einen lokal-kompakten Raum Y ist konservativ.*

Beweis. Es sei ν ein Maß aus $\mathscr{M}(Y, \varphi(X))$ und y ein Punkt aus $\varphi(X)$. Nach der Behauptung (b) des Hilfssatzes 3 gibt es eine offene Umgebung V von y in Y, für welche die zugehörige Abbildung $\varphi'_V: U \to V$ mit $U = \overset{-1}{\varphi}(V)$ eigentlich ist. Da ν auf $\varphi(X)$ konzentriert ist, so ist das von ν in V induzierte Maß ν_V auf $V \cap \varphi(X) = \varphi'_V(U)$ konzentriert ([6], p. 88, cor. 2). Aus dem Lokalisationsprinzip (Satz 1) folgt daher, daß es genügt, den Satz für eigentliche Abbildungen zu beweisen.

Die Abbildung φ werde also von nun an als eigentlich vorausgesetzt. Nach wie vor sei ν ein Maß aus $\mathscr{M}(Y, \varphi(X))$. Der Beweis der Darstellbarkeit von ν beruht auf dem folgenden allgemeinen Satz über die Fortsetzbarkeit positiver stetiger Linearformen[6]):

Es sei E ein geordneter lokal-konvexer Vektorraum (über $\mathbf{R}$*), P der konvexe Kegel aller Elemente $g \geqq 0$ aus E, E_0 ein linearer Unterraum von E und μ_0 eine auf E_0 definierte Linearform. Eine μ_0 auf ganz E fortsetzende, positive, stetige Linearform μ existiert dann und nur dann, wenn die Menge $\mu_0(E_0 \cap (V + P))$ reeller Zahlen für mindestens eine Umgebung V des Nullvektors $0 \in E$ nach unten beschränkt ist.* ($V + P$ bezeichnet hierbei die Menge aller Vektoren $v + p$ mit $v \in V$ und $p \in P$.)

In der nun folgenden Anwendung dieses Satzes ist E der geordnete Vektorraum $\mathscr{K}(X)$, $P = \mathscr{K}_+(X)$, E_0 die Menge aller Funktionen $f \circ \varphi$ mit $f \in \mathscr{K}(Y)$, die nach Hilfssatz 3 sämtlich in E liegen, und μ_0 die auf E_0 definierte Linearform $\mu_0(f \circ \varphi) = \int f\,d\nu$. Es genügt zu zeigen, daß μ_0 zu einer positiven Linearform μ auf E fortgesetzt werden kann. Dann ist nämlich μ ein positives Maß auf X mit $\varphi(\mu) = \nu$. Um den Fortsetzungssatz anwenden zu können, versehen wir noch E mit einer lokal-konvexen Topologie $\mathscr{T}$ und zeigen, daß μ sogar $\mathscr{T}$-stetig gewählt werden kann.

$\mathscr{T}$ sei die wie folgt definierte, sog. starke Topologie[7]) auf E: Für jede kompakte Menge K in X sei E_K die Menge aller Funktionen $g \in E$, die in $X - K$ gleich Null sind. E_K ist ein linearer Unterraum von E; er werde mit

[6]) Vgl. H. Bauer [3]. Unabhängig vom Verf. und etwa gleichzeitig wurde der Satz auch von I. Namioka [14] gefunden. Vgl. ferner G. Aumann [1].

[7]) Vgl. N. Bourbaki [5], p. 64, exercice 1.

der Topologie $\mathcal{T}_K$ der gleichmäßigen Konvergenz auf X versehen. Die Topologie auf E ist dann der induktive Limes der Topologien $\mathcal{T}_K$[8]). Da die Vereinigung aller Räume E_K gleich E ist, erhält man nach einer Bemerkung von N. Bourbaki [9], p. 62, ein Fundamentalsystem von Umgebungen des Nullvektors von E bezüglich $\mathcal{T}$, indem man alle Familien (V_K) betrachtet, in welchen jeder kompakten Menge $K \subset X$ eine symmetrische Umgebung V_K des Nullvektors von E_K bezüglich $\mathcal{T}_K$ zugeordnet ist, und die konvexe Hülle der Vereinigungsmenge aller V_K einer jeden solchen Familie bildet.

Die Existenz einer Fortsetzung μ ergibt sich nun wie folgt. Jeder kompakten Menge $K \subset X$ wird eine Funktion $f_K \in \mathcal{K}(Y)$ zugeordnet, und zwar derart, daß $0 \leq f_K \leq 1$ und $f_K(y) = 1$ ist für alle $y \in \varphi(K)$. Die Kompaktheit von $\varphi(K)$ sichert die Existenz einer solchen Funktion. Mittels f_K wird weiter jeder kompakten Menge $K \subset X$ eine Zahl $\alpha_K > 0$ zugeordnet, und zwar sei $\alpha_K = 1$, wenn $\int f_K d\nu = 0$ ist, und $\alpha_K = (\int f_K d\nu)^{-1}$, wenn $\int f_K d\nu > 0$ ist. Schließlich bezeichne V_K die Menge aller $g \in E_K$ mit $|g(x)| \leq \alpha_K$ für alle $x \in X$. Für jedes K ist dann V_K eine symmetrische Umgebung der Null in E_K bezüglich $\mathcal{T}_K$ und folglich die konvexe Hülle V der Vereinigungsmenge aller dieser V_K eine Umgebung der Null in E bezüglich $\mathcal{T}$. Nach dem allgemeinen Fortsetzungssatz genügt es zu zeigen, daß μ_0 auf der Menge $E_0 \cap (V + P)$ nach unten beschränkt ist. Jede Funktion aus E_0 ist von der Form $f \circ \varphi$ mit $f \in \mathcal{K}(Y)$. Zu jeder Funktion $f \circ \varphi \in E_0 \cap (V + P)$ gibt es endlich viele kompakte Mengen $K_1, \ldots, K_n$ in X, Funktionen $v_1, \ldots, v_n$ aus $V_{K_1}, \ldots, V_{K_n}$, reelle Zahlen $\lambda_1 \geq 0, \ldots, \lambda_n \geq 0$ mit $\sum \lambda_i = 1$ und eine Funktion $p \in P$ derart, daß

$$f \circ \varphi = \sum_{i=1}^{n} \lambda_i v_i + p \tag{3.1}$$

ist. Wegen $v_i \in V_{K_i}$ ist $v_i \geq -\alpha_{K_i}$ und $v_i(x) = 0$ für alle $x \in X - K_i$, also

$$v_i \geq -\alpha_{K_i} f_{K_i} \circ \varphi \qquad (i = 1, \ldots, n). \tag{3.2}$$

Wegen $p \geq 0$ folgt daher aus (3.1) und (3.2):

$$f \circ \varphi \geq -\sum_{i=1}^{n} \lambda_i \alpha_{K_i} f_{K_i} \circ \varphi. \tag{3.3}$$

Da das Maß ν auf $\varphi(X)$ konzentriert und somit die Menge $Y - \varphi(X)$ ν-lokal-vernachlässigbar ist, folgt weiter:

$$f \geq -\sum_{i=1}^{n} \lambda_i \alpha_{K_i} f_{K_i} \quad \text{lokal fast-überall bez. } \nu. \tag{3.4}$$

Integration bezüglich ν ergibt daher

$$\mu_0(f \circ \varphi) = \int f d\nu \geq -\sum_{i=1}^{n} \lambda_i \alpha_{K_i} \int f_{K_i} d\nu \geq -\sum_{i=1}^{n} \lambda_i = -1. \tag{3.5}$$

Also ist μ_0 auf $E_0 \cap (V + P)$ nach unten durch die Zahl -1 beschränkt.

Damit ist Satz 2 bewiesen.

[8]) Vgl. N. Bourbaki [9], chap. II, § 2, Nr. 4.

Bemerkungen. 1. Nach einer Mitteilung von Herrn Choquet findet sich der Satz 2 für den Spezialfall einer eigentlichen Abbildung von X auf Y, sowie unter einschränkenden Voraussetzungen über die lokal-kompakten Räume X und Y in einer noch unveröffentlichten Arbeit von G. Choquet und J. Deny über potentialtheoretische Fragen.

2. Für den Beweis des Satzes 2 hätte es genügt, den Vektorraum $E = \mathscr{K}(X)$ mit der feinsten lokal-konvexen Topologie zu versehen. Die von uns verwendete starke Topologie $\mathscr{T}$ ist aber in diesem Zusammenhang natürlicher, da bekanntlich die Radonschen Maße auf X gerade die $\mathscr{T}$-stetigen Linearformen auf E sind. (Vgl. das Zitat in [7].)

3.2. *Fortsetzung stetiger Abbildungen zu eigentlichen Abbildungen.* — Die Nr. 1.7 hat die Existenz stetiger, nicht-konservativer Abbildungen ergeben. Der folgende Satz 3 (zusammen mit Satz 2) zeigt, daß aber immerhin jede stetige Abbildung eines lokal-kompakten Raumes in einen ebensolchen Raum wenigstens die Restriktion einer stetigen konservativen, genauer sogar einer eigentlichen Abbildung ist. Wir behaupten nämlich:

Satz 3. *Es sei $\varphi: X \to Y$ eine stetige Abbildung eines lokal-kompakten Raumes X in einen lokal-kompakten Raum Y. Dann existiert ein lokal-kompakter Raum $\hat{X}$ mit folgenden Eigenschaften:*

(E_1) *Der Raum X ist ein Unterraum von $\hat{X}$ und liegt dicht in $\hat{X}$.*

(E_2) *Die Abbildung φ kann (auf genau eine Weise) zu einer stetigen Abbildung $\hat{\varphi}: \hat{X} \to Y$ fortgesetzt werden.*

(E_3) *Die Abbildung $\hat{\varphi}$ ist eigentlich.*

(E_4) *Für je zwei Punkte $x_1, x_2 \in \hat{X} - X$ mit $x_1 \neq x_2$ ist $\hat{\varphi}(x_1) \neq \hat{\varphi}(x_2)$.*

Zusatz. *Ein lokal-kompakter Raum $\hat{X}$ ist durch die Eigenschaften (E_1) — (E_4) bis auf Homöomorphien eindeutig bestimmt, die den Raum X punktweise festlassen.*

Der Zusatz berechtigt uns im folgenden von *dem* zu einer Abbildung φ gehörigen Raum $\hat{X}$ zu sprechen. Insbesondere ist die Identität der Räume X und $\hat{X}$ gleichbedeutend mit der Eigentlichkeit der Abbildung φ.

Da es sich bei Satz 3 und seinem Zusatz um einen rein topologischen Sachverhalt handelt, werden wir den Beweis hierfür erst in Nr. 3.3 erbringen und uns zunächst den maßtheoretischen Folgerungen zuwenden. An topologischen Eigenschaften von $\hat{X}$ erwähnen wir nur, daß der Raum X in $\hat{X}$ *offen* ist. Es ist nämlich X als lokal-kompakter Unterraum von $\hat{X}$ von der Form $X = \hat{F} \cap \hat{G}$, wobei $\hat{F}$ in $\hat{X}$ abgeschlossen und $\hat{G}$ in $\hat{X}$ offen ist. Da X in $\hat{X}$ dicht liegt, muß $\hat{F} = \hat{X}$ und somit $X = \hat{G}$ sein.

Zunächst bemerken wir, daß für eine stetige Abbildung $\varphi: X \to Y$ zwar nicht jedes Maß $\nu \in \mathscr{M}(Y, \varphi(X))$ φ-darstellbar zu sein braucht, aber aus Satz 2 folgt:

Jedes auf $\varphi(X)$ konzentrierte positive Maß auf Y ist darstellbar bezüglich der Abbildung $\hat{\varphi}: \hat{X} \to Y$.

Weiter können wir eine neue Interpretation des bereits beobachteten Phänomens geben, wonach nur für gewisse Maße $\mu \in \mathcal{M}(X)$ das Bild $\varphi(\mu) \in \mathcal{M}(Y)$ definiert ist:

Satz 4. *Für jede stetige Abbildung $\varphi: X \to Y$ eines lokal-kompakten Raumes X in einen lokal-kompakten Raum Y und jedes Maß $\mu \in \mathcal{M}(X)$ sind folgende Aussagen gleichwertig:*

(a) *Das Bildmaß $\varphi(\mu)$ ist definiert.*

(b) *Das Maß μ wird durch ein Maß $\tilde{\mu} \in \mathcal{M}(\tilde{X})$ in X induziert.*

Beweis. Im Hinblick auf das in Nr. 1.5 Gesagte läßt sich die Behauptung auch so aussprechen: Die Abbildung φ ist dann und nur dann μ-eigentlich, wenn die kanonische Injektion $j: X \to \tilde{X}$ μ-eigentlich ist. Wir beweisen die Behauptung in dieser Formulierung.

Sei also zunächst φ μ-eigentlich. Dann ist für jede in $\tilde{X}$ kompakte Menge $\tilde{K}$ und deren kompaktes Bild $L = \tilde{\varphi}(\tilde{K})$ die Urbildmenge $\overset{-1}{\varphi}(L)$ im wesentlichen μ-integrierbar. Nun ist aber $X \cap \tilde{K}$ abgeschlossen in X und somit μ-meßbar; ferner ist $X \cap \tilde{K} \subset \overset{-1}{\varphi}(L)$. Nach Nr. 1.3 ist daher mit $\overset{-1}{\varphi}(L)$ auch $X \cap \tilde{K} = \overset{-1}{j}(\tilde{K})$ im wesentlichen μ-integrierbar, also die Injektion j μ-eigentlich. — Nun sei umgekehrt j μ-eigentlich. Die eigentliche Abbildung $\tilde{\varphi}$ ist für jedes Maß $\tilde{\nu} \in \mathcal{M}(\tilde{X})$ $\tilde{\nu}$-eigentlich. Nach dem Transitivitätsgesetz für Bildmaße ist daher auch $\varphi = \tilde{\varphi} \circ j$ μ-eigentlich.

Korollar. *Ein Maß $\nu \in \mathcal{M}(Y)$ ist dann und nur dann φ-darstellbar, wenn ein auf X konzentriertes Maß $\tilde{\mu} \in \mathcal{M}(\tilde{X})$ existiert mit $\tilde{\varphi}(\tilde{\mu}) = \nu$.*

Beweis. Gilt $\varphi(\mu) = \nu$ für ein $\mu \in \mathcal{M}(X)$, so folgt die Existenz eines solchen Maßes $\tilde{\mu}$ aus Satz 4 und dem Transitivitätsgesetz für Bildmaße. Dann ist nämlich $\tilde{\mu} = j(\mu)$ definiert für die kanonische Injektion $j: X \to \tilde{X}$ und ein Maß aus $\mathcal{M}(\tilde{X}, X)$; wegen $\varphi = \tilde{\varphi} \circ j$ ist ferner $\nu = \tilde{\varphi}(\tilde{\mu})$. — Ist umgekehrt $\tilde{\mu}$ ein Maß aus $\mathcal{M}(\tilde{X}, X)$ mit $\varphi(\tilde{\mu}) = \nu$ und μ das von $\tilde{\mu}$ in X induzierte Maß, so ist abermals nach Satz 4 $\varphi(\mu)$ definiert und $\varphi(\mu) = \nu$ wegen $\varphi = \tilde{\varphi} \circ j$ und $j(\mu) = \tilde{\mu}$.

3.3. *Beweis des Satzes* 3. — Zur Vorbereitung der nachfolgenden Konstruktion des Raumes $\tilde{X}$ beweisen wir zunächst:

Hilfssatz 4. *Für eine stetige Abbildung $\varphi: X \to Y$ eines lokal-kompakten Raumes X in einen lokal-kompakten Raum Y sei $\tilde{X}$ ein lokal-kompakter Raum mit den Eigenschaften (E_1)—(E_4). Dann wird der Unterraum $\tilde{X} - X$ von $\tilde{X}$ durch die stetige Fortsetzung $\tilde{\varphi}: \tilde{X} \to Y$ von φ homöomorph auf den Unterraum Y_0 von Y abgebildet, der aus allen Punkten $y \in Y$ besteht mit der Eigenschaft, daß $\overset{-1}{\varphi}(V)$ für keine Umgebung V von y in Y kompakt in X ist.*

Beweis. Wir setzen $Y_0' = \tilde{\varphi}(\tilde{X} - X)$ und bezeichnen mit $\psi: \tilde{X} - X \to Y_0'$ die durch $\tilde{\varphi}$ vermittelte stetige Abbildung von $\tilde{X} - X$ auf Y_0'. Der Unterraum $\tilde{X} - X$ von $\tilde{X}$ ist abgeschlossen, insbesondere also lokal-kompakt; die Abbildung $\tilde{\varphi}$ ist eigentlich und daher abgeschlossen. Also ist Y_0' in Y abgeschlossen

und somit selbst lokal-kompakt. Jede in Y_0' kompakte Menge K ist kompakt in Y und daher ist $\overset{-1}{\psi}(K) = (\tilde{X} - X) \cap \tilde{\varphi}^{-1}(K)$ kompakt in $\tilde{X} - X$. Die Abbildung ψ ist also ebenfalls eigentlich und folglich abgeschlossen. Da ferner ψ nach (E_4) eineindeutig ist, wird der Raum $\tilde{X} - X$ durch ψ homöomorph auf Y_0' abgebildet.

Nach diesem ersten Beweisschritt fehlt nur noch der Beweis der Gleichheit $Y_0' = Y_0$. Wäre für eine Umgebung V eines Punktes $y \in Y_0'$ die Urbildmenge $\overset{-1}{\varphi}(V)$ kompakt in X und damit auch in $\tilde{X}$, so gäbe es wegen der Stetigkeit von $\tilde{\varphi}$ eine zu $\overset{-1}{\varphi}(V)$ fremde Umgebung $\tilde{U}$ von $x = \overset{-1}{\psi}(y)$ in $\tilde{X}$ mit $\tilde{\varphi}(\tilde{U}) \subset V$. Da X in $\tilde{X}$ dicht liegt, gäbe es einen Punkt $x' \in X \cap \tilde{U}$. Für diesen wäre dann $\tilde{\varphi}(x') = \varphi(x') \in V$, also $x' \in \overset{-1}{\varphi}(V)$ im Widerspruch zu $\tilde{U} \cap \overset{-1}{\varphi}(V) = \theta$. Es ist also $\overset{-1}{\varphi}(V)$ für keine Umgebung V in Y eines Punktes $y \in Y_0'$ kompakt in X, d. h. es ist $Y_0' \subset Y_0$. — Da Y_0' in Y abgeschlossen ist, gibt es zu jedem Punkt $y \in Y - Y_0'$ eine zu Y_0' fremde, kompakte Umgebung V von y in Y. Wegen $V \cap Y_0' = \theta$ ist $\tilde{\varphi}^{-1}(V) \subset X$, also $\overset{-1}{\varphi}(V) = \tilde{\varphi}^{-1}(V)$ kompakt in X und somit $y \notin Y_0$. Also ist auch $Y - Y_0' \subset Y - Y_0$ und damit die Gleichheit $Y_0' = Y_0$ bewiesen.

Nun erst erbringen wir den

Beweis des Satzes 3. Wir bezeichnen hierzu wieder mit Y_0 den in Hilfssatz 4 definierten Unterraum von Y bezüglich der gegebenen Abbildung φ und bemerken zunächst, daß Y_0 in Y abgeschlossen ist. Zu jeder Umgebung V von $y \in \overline{Y}_0$ in Y existiert nämlich ein $y_0 \in V \cap Y_0$ derart, daß V auch Umgebung von y_0 ist. Daher muß $y \in Y_0$ sein.

Es sei weiter $\psi: X_0 \to Y_0$ eine Homöomorphie eines zu X punktfremden Raumes X_0 auf den Raum Y_0. Wir setzen

$$\tilde{X}_0 = X \cup X_0 \tag{3.6}$$

und definieren die folgende Abbildung $\tilde{\varphi}_0: \tilde{X}_0 \to Y$:

$$\tilde{\varphi}_0(x) = \begin{cases} \varphi(x), \text{ für jedes } x \in X\,; \\ \psi(x), \text{ für jedes } x \in X_0\,. \end{cases} \tag{3.7}$$

Ferner sei $\tilde{\mathfrak{B}}_0$ das System aller Mengen $\tilde{U} \subset \tilde{X}_0$ der folgenden Gestalt:

$$\tilde{U} = [U \cup \overset{-1}{\varphi}(V) \cup \overset{-1}{\psi}(V \cap Y_0)] \cap (\tilde{X}_0 - K)\,, \tag{3.8}$$

wobei U bzw. K eine in X offene bzw. kompakte Menge und V eine in Y offene Menge ist.

Wir überlassen es dem Leser nachzuprüfen, daß $\tilde{\mathfrak{B}}_0$ mit je endlich vielen Mengen auch deren Durchschnitt und insbesondere auch die leere Menge enthält. Wir versehen $\tilde{X}_0$ mit der Topologie, deren System aller offenen Mengen das System $\tilde{\mathfrak{O}}_0$ aller Vereinigungen von Mengen aus $\tilde{\mathfrak{B}}_0$ ist. Aus der speziellen Gestalt der Mengen von $\tilde{\mathfrak{O}}_0$ folgt, daß X ein offener Unterraum von $\tilde{X}_0$ ist. Ferner liegt X in $\tilde{X}_0$ dicht. Hierzu genügt es zu zeigen, daß aus $\tilde{U} \cap X = \theta$ folgt $\tilde{U} = \theta$ für jede Menge $\tilde{U}$ der Gestalt (3.8). Nun ist

$$\tilde{U} \cap X = [U \cup \overset{-1}{\varphi}(V)] \cap (X - K)\,,$$

wegen $\tilde{U} \cap X = \theta$ also $\overset{-1}{\varphi}(V) \subset U \cup \overset{-1}{\varphi}(V) \subset K$. Hieraus folgt $V \cap Y_0 = \theta$, da jeder Punkt $y \in V$ eine kompakte Umgebung $W \subset V$ besitzt mit $\overset{-1}{\varphi}(W) \subset \overset{-1}{\varphi}(V) \subset K$, also mit kompaktem Urbild. Dann aber ist $U \cup \overset{-1}{\varphi}(V) \cup \overset{-1}{\psi}(V \cap Y_0) = U \cup \overset{-1}{\varphi}(V)$ fremd zu K und somit tatsächlich $\tilde{U} = \theta$. Die Abbildung $\tilde{\varphi}_0 : \tilde{X}_0 \to Y$ ist eine stetige Fortsetzung von φ auf $\tilde{X}_0$, denn nach (3.7) gilt für jede in Y offene Menge V:

$$\tilde{\varphi}_0^{-1}(V) = \overset{-1}{\varphi}(V) \cup \overset{-1}{\psi}(V \cap Y_0) .$$

Schließlich ist der Raum $\tilde{X}_0$ separiert: Es seien hierzu x_1 und x_2 zwei verschiedene Punkte aus $\tilde{X}_0$. Sie besitzen fremde Umgebungen in $\tilde{X}_0$, falls sie beide in X liegen, da X separiert und offener Unterraum von $\tilde{X}_0$ ist, oder falls genau einer von ihnen in X liegt, da X lokal-kompakt und nach (3.8) die Komplemente (in $\tilde{X}_0$) kompakter Teilmengen von X offen in $\tilde{X}_0$ sind. Liegen jedoch beide Punkte in X_0, so ist nach (3.7) $\tilde{\varphi}_0(x_1) \neq \tilde{\varphi}_0(x_2)$ und die Existenz fremder Umgebungen folgt aus der Separiertheit von Y und der Stetigkeit von $\tilde{\varphi}_0$.

Somit ist $\tilde{X}_0$ ein Raum mit den gewünschten Eigenschaften, wenn gezeigt werden kann, daß das Urbild $\tilde{\varphi}_0^{-1}(L)$ einer jeden in Y kompakten Menge L kompakt in $\tilde{X}_0$ ist. Wegen der lokalen Kompaktheit von Y und der Stetigkeit von $\tilde{\varphi}_0$ ist dann nämlich $\tilde{X}_0$ lokal-kompakt und $\tilde{\varphi}_0$ eigentlich. Also sei L kompakt in Y. Dann ist

$$\tilde{\varphi}_0^{-1}(L) = \overset{-1}{\varphi}(L) \cup \overset{-1}{\psi}(L \cap Y_0) . \tag{3.9}$$

Es sei $\mathfrak{U}$ ein Ultrafilter in $\tilde{\varphi}_0^{-1}(L)$; zu zeigen ist dessen Konvergenz in $\tilde{\varphi}_0^{-1}(L)$. Es genügt sogar, nur die Konvergenz der Filterbasis $\mathfrak{U}$ in $\tilde{X}_0$ zu beweisen, da $\tilde{\varphi}_0^{-1}(L)$ wegen der Stetigkeit von $\tilde{\varphi}_0$ abgeschlossen ist. Aus (3.9) folgt, daß entweder $\overset{-1}{\psi}(L \cap Y_0)$ oder $\overset{-1}{\varphi}(L)$ ein Element von $\mathfrak{U}$ ist. Da ψ eine Homöomorphie, L kompakt und Y_0 in Y abgeschlossen ist, so ist $\overset{-1}{\psi}(L \cap Y_0)$ kompakt in $\tilde{X}_0$. Also konvergiert im ersten Falle, wo diese Menge zu $\mathfrak{U}$ gehört, der Ultrafilter $\mathfrak{U}$ gegen einen Punkt von $\overset{-1}{\psi}(L \cap Y_0)$. Zu behandeln ist daher nur noch der zweite Fall: $\overset{-1}{\varphi}(L) \in \mathfrak{U}$. Hier induziert $\mathfrak{U}$ einen Ultrafilter $\mathfrak{U}'$ in $L' = \overset{-1}{\varphi}(L)$ mit $\mathfrak{U}' \subset \mathfrak{U}$; $\mathfrak{V} = \varphi(\mathfrak{U}')$ ist die Basis eines Ultrafilters in L. Da L kompakt ist, konvergiert somit $\mathfrak{V}$ gegen einen Punkt $y \in L$. Wiederum sind nun zwei Fälle zu unterscheiden. 1. Fall: Es ist y ein Punkt aus Y_0, also $y \in L \cap Y_0$. Konvergiert $\mathfrak{U}$ gegen einen Punkt aus X, so ist nichts mehr zu beweisen. Konvergiert aber $\mathfrak{U}$ nicht in X, so konvergiert $\mathfrak{U}$ in $\tilde{X}_0$ gegen $x = \overset{-1}{\psi}(y)$. In der Tat: Es sei

$$\tilde{U} = [U \cup \overset{-1}{\varphi}(V) \cup \overset{-1}{\psi}(V \cap Y_0)] \cap (\tilde{X}_0 - K)$$

eine Menge aus $\tilde{\mathfrak{B}}_0$ (d. h. U offen in X, K kompakt in X, V offen in Y) mit $x \in \tilde{U}$, also mit $y \in V$. Da $\mathfrak{V}$ gegen y konvergiert, gibt es eine Menge $W_1 \in \mathfrak{U}' \subset \mathfrak{U}$ mit $\varphi(W_1) \subset V$, also mit $W_1 \subset \overset{-1}{\varphi}(V)$. Da $\mathfrak{U}$ nicht in X konvergiert, gibt es weiter zur kompakten Menge K eine Menge $W_2 \in \mathfrak{U}$ mit $W_2 \subset \tilde{X}_0 - K$. Für die zu $\mathfrak{U}$ gehörige Menge $W = W_1 \cap W_2$ gilt dann offenbar: $W \subset \tilde{U}$. Nun bilden aber die

x enthaltenden Mengen $\tilde{U} \in \tilde{\mathfrak{B}}_0$ ein fundamentales Umgebungssystem von x. Daher konvergiert $\mathfrak{U}$ gegen x. — 2. Fall: Es ist $y \notin Y_0$. Dann gibt es eine kompakte Umgebung V von y in Y mit in X und $\tilde{X}_0$ kompaktem Urbild $\overset{-1}{\varphi}(V)$. Da $\mathfrak{B}$ gegen y konvergiert, gibt es eine Menge $W \in \mathfrak{U}'$ mit $\varphi(W) \subset V \cap L$, also mit $W \subset \overset{-1}{\varphi}(V \cap L)$. Also enthält $\mathfrak{U}'$ und damit $\mathfrak{U}$ die zufolge der Stetigkeit von φ und der Kompaktheit von $\overset{-1}{\varphi}(V)$ kompakte Menge $\overset{-1}{\varphi}(V \cap L)$. Dann aber konvergiert $\mathfrak{U}$ gegen einen Punkt dieser Menge, womit unsere Fallunterscheidung zu Ende geführt ist.

Der konstruierte Raum $\tilde{X}_0$ ist also ein lokal-kompakter Raum mit den Eigenschaften (E_1) — (E_4).

Abschließend führen wir noch den

Beweis des Zusatzes. Wir bezeichnen wieder mit $\mathscr{K}(X)$ bzw. $\mathscr{K}(Y)$ den Vektorraum aller stetigen reellen Funktionen mit kompaktem Träger auf X bzw. Y; ferner sei $\mathscr{A}$ die Menge aller Funktionen $g = k + f \circ \varphi$ mit $k \in \mathscr{K}(X)$ und $f \in \mathscr{K}(Y)$. Dann ist $\mathscr{A}$ bezüglich der üblichen Operationen eine Algebra über dem Körper $\mathbf{R}$. Ist nun $\tilde{X}$ ein lokal-kompakter Raum mit den Eigenschaften (E_1) — (E_4), so kann jede Funktion $g = k + f \circ \varphi$ aus $\mathscr{A}$ auf genau eine Weise zu einer stetigen reellen Funktion $\tilde{g}$ auf $\tilde{X}$ fortgesetzt werden: es ist $\tilde{g} = \tilde{k} + f \circ \tilde{\varphi}$, wenn hierbei $\tilde{k}$ diejenige Fortsetzung von k bezeichnet, welche auf $\tilde{X} - X$ gleich Null ist. Da die Abbildung $\tilde{\varphi}$ eigentlich und $\tilde{k}$ ein Element von $\mathscr{K}(\tilde{X})$ ist, gilt $\tilde{g} \in \mathscr{K}(\tilde{X})$ für alle $g \in \mathscr{A}$. Die Menge $\tilde{\mathscr{A}}$ aller Fortsetzungen $\tilde{g}$ von Funktionen $g \in \mathscr{A}$ ist dann ebenfalls eine Algebra. $\tilde{\mathscr{A}}$ trennt die Punkte von $\tilde{X}$, d. h. zu je zwei Punkten $x_1 \neq x_2$ aus $\tilde{X}$ gibt es ein $\tilde{g} \in \tilde{\mathscr{A}}$ mit $\tilde{g}(x_1) \neq \tilde{g}(x_2)$. In der Tat: liegen x_1 und x_2 in $\tilde{X} - X$, so gilt $\tilde{\varphi}(x_1) \neq \tilde{\varphi}(x_2)$ nach (E_4); also gibt es ein $f \in \mathscr{K}(Y)$ mit $f(\tilde{\varphi}(x_1)) \neq f(\tilde{\varphi}(x_2))$; es ist aber $f \circ \tilde{\varphi} \in \tilde{\mathscr{A}}$. Liegen x_1 und x_2 nicht beide in $\tilde{X} - X$, so gilt $\tilde{k}(x_1) \neq \tilde{k}(x_2)$ für ein geeignetes $k \in \mathscr{K}(X)$. Eine analoge Überlegung zeigt, daß es zu jedem $x \in \tilde{X}$ mindestens ein $\tilde{g} \in \tilde{\mathscr{A}}$ gibt mit $\tilde{g}(x) \neq 0$. Nach der von H. Nakano [13] stammenden Variante des Approximationssatzes von Stone-Weierstrass (vgl. auch [16], S. 72) kann daher jede auf $\tilde{X}$ stetige, im Unendlichen verschwindende, reelle Funktion[9]) gleichmäßig auf $\tilde{X}$ durch Funktionen aus $\tilde{\mathscr{A}}$ approximiert werden.

Jetzt seien $\tilde{X}_1$ und $\tilde{X}_2$ zwei lokal-kompakte Räume mit den Eigenschaften (E_1) — (E_4). $\mathscr{C}_0(\tilde{X}_i)$ bezeichne die Algebra aller auf $\tilde{X}_i$ stetigen reellen Funktionen, die im Unendlichen verschwinden ($i = 1, 2$). Der erste Teil unseres Beweises zeigt, daß eine auf $\tilde{X}_1$ definierte Funktion $\tilde{g}_1$ dann und nur dann in $\mathscr{C}_0(\tilde{X}_1)$ liegt, wenn ihre Restriktion g auf X zu einer Funktion $\tilde{g}_2 \in \mathscr{C}_0(\tilde{X}_2)$ fortgesetzt werden kann. Für jedes $\tilde{g}_1 \in \mathscr{C}_0(\tilde{X}_1)$ ist diese Funktion $\tilde{g}_2 \in \mathscr{C}_0(\tilde{X}_2)$ eindeutig bestimmt. Die Abbildung $\tilde{g}_1 \to \tilde{g}_2$ ist ein Isomorphismus $\Psi: \mathscr{C}_0(\tilde{X}_1) \to$

---

[9]) Bekanntlich heißt eine auf einem lokal-kompakten Raum T definierte, stetige, reelle Funktion f im Unendlichen verschwindend, wenn für jede reelle Zahl $\alpha > 0$ die Menge aller $x \in T$ mit $|f(x)| \geq \alpha$ kompakt ist. Für kompaktes T hat offenbar jede stetige, reelle Funktion diese Eigenschaft.

$\mathscr{C}_0(\bar{X}_2)$ der Algebra $\mathscr{C}_0(\bar{X}_1)$ auf die Algebra $\mathscr{C}_0(\bar{X}_2)$. Nach bekannten Sätzen (vgl. etwa [15], Satz 2, und [7], p. 100, théorème 4) existiert dann eine Homöomorphie $\psi : \bar{X}_2 \to \bar{X}_1$ von $\bar{X}_2$ auf $\bar{X}_1$ derart, daß für alle $\bar{g}_1 \in \mathscr{C}_0(\bar{X}_1)$ gilt:

$$\Psi(\bar{g}_1) = \bar{g}_1 \circ \psi .$$

Für jede Funktion $k \in \mathscr{K}(X)$ und deren stetige Fortsetzungen $\bar{k}_1$ bzw. $\bar{k}_2$ auf $\bar{X}_1$ bzw. $\bar{X}_2$ ist daher $\bar{k}_2 = \bar{k}_1 \circ \psi$, also insbesondere $k(x) = \bar{k}_1(\psi(x))$ für jedes $x \in X$. Da $\bar{k}_1$ in $\bar{X}_1 - X$ gleich Null ist, muß dann $\psi(x) \in X$ und, da $\mathscr{K}(X)$ die Punkte von X trennt, schließlich $\psi(x) = x$ für jedes $x \in X$ sein. Also ist ψ eine den Raum X punktweise festlassende Homöomorphie von $\bar{X}_2$ auf $\bar{X}_1$.

§ 4. Im Unendlichen abzählbare Räume

Wir kehren für einen Augenblick zurück zu dem in Nr. 1.7 konstruierten Beispiel einer nicht-konservativen, stetigen Abbildung $\varphi : X \to Y$. Das dort bewiesene Resultat, durch welches die φ-darstellbaren Maße $\nu \in \mathscr{M}(Y)$ charakterisiert wurden, kann auch folgendermaßen ausgesprochen werden: In diesem Beispiel ist ein Maß $\nu \in \mathscr{M}(Y)$ dann und nur dann φ-darstellbar, wenn es von einer Menge $M \subset \varphi(X)$ getragen wird, deren Urbild $\overset{-1}{\varphi}(M)$ in der Vereinigung abzählbar vieler kompakter Teilmengen von X enthalten ist. In der Tat werden hierdurch im Beispiel gerade die abzählbaren Teilmengen M von Y gekennzeichnet. Wir werden nun zunächst zeigen, daß die Aussage „dann" allgemein richtig ist.

Satz 5. *Es sei $\varphi : X \to Y$ eine stetige Abbildung eines lokal-kompakten Raumes X in einen lokal-kompakten Raum Y. Dann ist jedes Maß $\nu \in \mathscr{M}(Y)$ φ-darstellbar, welches von einer Menge $M \subset \varphi(X)$ getragen wird, deren Urbild $\overset{-1}{\varphi}(M)$ in der Vereinigung abzählbarer vieler kompakter Teilmengen von X enthalten ist.*

Beweis. Nach Voraussetzung gibt es eine Folge $(K_n)_{n=1,2,\dots}$ kompakter Mengen in X mit $\overset{-1}{\varphi}(M) \subset \bigcup K_n$. Ohne Beschränkung der Allgemeinheit kann $K_n \subset K_{n+1}$ für alle $n = 1, 2, \dots$ angenommen werden. Jede der Bildmengen $L_n = \varphi(K_n)$ ist kompakt in Y; es gilt $L_n \subset L_{n+1}$ für alle n und $M \subset L = \bigcup L_n$, letzteres wegen $M \subset \varphi(X)$. Die Menge L ist als Vereinigung einer Folge kompakter und daher ν-meßbarer Mengen selbst ν-meßbar. Also wird ν auch von der Menge L getragen. Für jedes n bezeichne nun ν_n das aus ν durch Tilgung der Masse außerhalb L_n entstehende Maß: $\nu_n = \chi_{L_n} \nu$. Da ν auf L konzentriert ist und wegen $L_n \subset L_{n+1} \subset L$, gilt:

$$\nu_n \leqq \nu_{n+1} \leqq \nu \qquad (n = 1, 2, \dots). \tag{4.1}$$

Ferner ist

$$\lim \nu_n = \nu , \tag{4.2}$$

d. h. die Folge (ν_n) konvergiert vage gegen ν. In der Tat: für beliebiges $f \in \mathscr{K}_+(Y)$ ist jede der Funktionen $f_n = \chi_{L_n} f$ ν-integrierbar; wegen $f_n \leqq f_{n+1} \leqq$ $\leqq f \chi_L = \sup f_n$ gilt somit nach dem Lebesgueschen Konvergenzsatz $\lim \int f_n d\nu$ $= \int f \chi_L d\nu = \int f d(\chi_L \nu)$. Hieraus folgt (4.2), da $\nu_n(f) = \int f_n d\nu$ und $\nu = \chi_L \nu$ ist.

Nach dieser Vorbereitung behaupten wir die Existenz einer Folge $(\mu_n)_{n=1,2,\ldots}$ von Maßen $\mu_n \in \mathcal{M}(X)$ mit folgenden drei Eigenschaften: (a) μ_n ist auf K_n konzentriert; (b) $\varphi(\mu_n) = \nu_n$; (c) $\mu_n \leqq \mu_{n+1}$ $(n = 1, 2, \ldots)$. Hierzu bezeichne $j_n : K_n \to X$ bzw. $j'_n : L_n \to Y$ die kanonische Injektion von K_n in X bzw. L_n in Y, $\varphi'_n : K_n \to L_n$ diejenige Abbildung, die auf K_n mit φ übereinstimmt, und ν_{L_n} das durch ν in L_n induzierte Maß. Dann ist jede der Abbildungen φ'_n eigentlich und bildet K_n auf L_n ab, ferner ist $\varphi \circ j_n = j'_n \circ \varphi'_n$ und $j'_n(\nu_{L_n}) = \nu_n$. Wegen der beiden ersten Eigenschaften von φ'_1 existiert nach Satz 2 zu ν_{L_1} ein Maß $\mu'_1 \in \mathcal{M}(K_1)$ mit $\varphi'_1(\mu'_1) = \nu_{L_1}$. Da die Abbildung $j_1 : K_1 \to X$ eigentlich ist, so ist $\mu_1 = j(\mu'_1)$ definiert und ein auf K_1 konzentriertes Maß aus $\mathcal{M}(X)$. Das Transitivitätsgesetz für Bildmaße ergibt dann bei Beachtung von $\varphi \circ j_1 = j'_1 \circ \varphi'_1$ und $j'_1(\nu_{L_1}) = \nu_1$ die Gleichung $\varphi(\mu_1) = \nu_1$. Ausgehend von μ_1 konstruieren wir die weiteren Elemente der Folge (μ_n) durch vollständige Induktion. Wir nehmen hierzu an, daß $\mu_1, \ldots, \mu_n$ für ein $n \geqq 1$ bereits konstruiert sind. Nach (4.1) ist $\tau = \nu_{n+1} - \nu_n$ ein auf L_{n+1} konzentriertes Maß aus $\mathcal{M}(Y)$; für das hierdurch in L_{n+1} induzierte Maß $\tau_{L_{n+1}}$ gilt daher $j'_{n+1}(\tau_{L_{n+1}}) = \tau$. Wiederum nach Satz 2 gibt es ein Maß $\sigma' \in \mathcal{M}(K_{n+1})$ mit $\varphi'_{n+1}(\sigma') = \tau_{L_{n+1}}$. Eine Wiederholung der soeben angestellten Überlegungen ergibt, daß $\sigma = j_{n+1}(\sigma')$ ein Maß aus $\mathcal{M}(X, K_{n+1})$ ist mit $\varphi(\sigma) = \tau$. Setzen wir daher $\mu_{n+1} = \mu_n + \sigma$, so ist μ_{n+1} ein auf K_{n+1} konzentriertes Maß aus $\mathcal{M}(X)$ mit $\mu_n \leqq \mu_{n+1}$; nach Hilfssatz 1 ist ferner $\varphi(\mu_{n+1}) = \varphi(\mu_n) + \varphi(\sigma) = \nu_n + \tau = \nu_{n+1}$. Damit ist die Existenz einer Folge (μ_n) mit den Eigenschaften (a)—(c) gesichert.

Wir wollen jetzt zeigen, daß die Folge (μ_n) vage konvergiert[10]). Wegen der Eigenschaft (c) genügt es zu beweisen, daß die Folge $(\int g\,d\mu_n)$ für jedes $g \in \mathcal{K}_+(X)$ nach oben beschränkt ist. Zu $g \in \mathcal{K}_+(X)$ gibt es aber, wie im Beweis des Satzes 1 gezeigt wurde, eine Funktion $f \in \mathcal{K}_+(Y)$ mit $0 \leqq g \leqq f \circ \varphi$. Aus (b) und (4.1) folgt daher

$$\int g\,d\mu_n \leqq \int f \circ \varphi\,d\mu_n = \int f\,d\nu_n \leqq \int f\,d\nu \qquad (n = 1, 2, \ldots). \tag{4.3}$$

Damit ist die Existenz des vagen Limes $\mu = \lim \mu_n$ bewiesen. Wir behaupten weiter, daß $\varphi(\mu) = \nu$ ist. Dazu genügt es zu zeigen, daß für jedes $f \in \mathcal{K}_+(Y)$ die Funktion $f \circ \varphi$ μ-integrierbar und $\int f \circ \varphi\,d\mu = \int f\,d\nu$ ist. Zunächst ergibt sich genauso wie soeben die Gültigkeit von (4.3) für jedes $g \in \mathcal{K}_+(X)$ mit $0 \leqq g \leqq f \circ \varphi$. Hieraus folgt $\int g\,d\mu \leqq \int f\,d\nu$ durch Grenzübergang. Also ist die Funktion $f \circ \varphi$ μ-integrierbar und $\int f \circ \varphi\,d\mu \leqq \int f\,d\nu$ (vgl. Nr. 1.2). Wegen $\mu_n \leqq \mu$ und $\varphi(\mu_n) = \nu_n$ ist weiter $\int f\,d\nu_n = \int f \circ \varphi\,d\mu_n \leqq \int f \circ \varphi\,d\mu$, woraus bei Beachtung von (4.2) folgt: $\int f\,d\nu \leqq \int f \circ \varphi\,d\mu$ und damit $\int f \circ \varphi\,d\mu = \int f\,d\nu$. Wie behauptet ist also $\varphi(\mu) = \nu$.

Jetzt sei an folgende bekannte Begriffsbildung erinnert:

Definition 4. *Ein lokal-kompakter Raum T heißt* ***abzählbar im Unendlichen****, wenn eine Folge $(K_n)_{n=1,2,\ldots}$ in T kompakter Mengen existiert mit $T = \bigcup K_n$.*

---

[10]) Der Rest des Beweises ähnelt einem Abschnitt des Beweises von Satz 1.

Dann ergibt sich mühelos das folgende Korollar zu Satz 5:

Korollar. *Jede stetige Abbildung $\varphi : X \to Y$ eines im Unendlichen abzählbaren, lokal-kompakten Raumes X in einen lokal-kompakten Raum Y ist konservativ.*

Unter Verwendung des Lokalisationsprinzips kann dieses Korollar verschärft werden zum

Satz 6. *Es sei $\varphi : X \to Y$ eine stetige Abbildung eines lokal-kompakten Raumes X in einen lokal-kompakten Raum Y. Besitzt dann jeder Punkt $y \in \varphi(X)$ eine offene Umgebung V in Y, deren Urbild $\overset{-1}{\varphi}(V)$, aufgefaßt als Unterraum von X, im Unendlichen abzählbar ist, so ist φ konservativ.*

Beweis. Die Behauptung folgt aus dem Korollar zu Satz 5 und aus Satz 1. Man hat nur wie zu Beginn des Beweises von Satz 2 zu bemerken, daß aus $\nu \in \mathscr{M}(Y, \varphi(X))$ folgt $\nu_V \in \mathscr{M}(V, \varphi(X) \cap V)$ für jede offene Menge $V \subset Y$.

Bemerkungen. 1. Ist $\varphi : X \to Y$ eine stetige Abbildung eines im Unendlichen abzählbaren, lokal-kompakten Raumes X in einen lokal-kompakten Raum Y, so ist offenbar das Urbild $\overset{-1}{\varphi}(K)$ einer jeden in Y kompakten Menge K darstellbar als Vereinigung abzählbar vieler kompakter Mengen in X. Es wäre interessant zu wissen, ob diese Bedingung allein schon impliziert, daß die Abbildung konservativ ist. Dann wären nämlich Satz 2 und das Korollar zu Satz 5 Folgerungen aus einem allgemeineren Satz.

2. Auf Grund der engen Beziehungen zwischen den im Unendlichen abzählbaren, lokal-kompakten Räumen und den parakompakten, lokal-kompakten Räumen[11]), könnte man vermuten, daß jede stetige Abbildung eines parakompakten, lokal-kompakten Raumes X in einen lokal-kompakten Raum Y konservativ ist. Daß dies jedoch nicht der Fall ist, zeigt das Beispiel in Nr. 1.7; der dort betrachtete Raum X ist parakompakt und lokal-kompakt.

§ 5. Abbildungen mit lokalem Schnitt

Zunächst seien X und Y beliebige topologische Räume und $\varphi : X \to Y$ eine stetige Abbildung. Dann erweist es sich häufig als zweckmäßig, das Tripel (X, Y, φ) als einen verallgemeinerten Faserraum aufzufassen, dessen Fasern im allgemeinen zueinander nicht homöomorph sind[12]).

Fasern der Abbildung φ werden hierbei die nicht-leeren Urbilder einpunktiger Teilmengen von Y, also alle Mengen $\overset{-1}{\varphi}(y)$ mit $y \in \varphi(X)$ genannt. Vermöge dieser Definition ist keine Faser leer, je zwei verschiedene Fasern sind zueinander fremd und jeder Punkt $x \in X$ liegt in genau einer Faser $F = \overset{-1}{\varphi}(\varphi(x))$, genannt die Faser durch x oder über $\varphi(x)$. Die Fasern von φ definieren also in X eine Klasseneinteilung.

Der aus der Theorie der Faserräume geläufige Begriff des (lokalen) Schnittes läßt sich auf unsere allgemeinere Situation übertragen:

[11]) Vgl. N. Bourbaki [7], p. 107, théorème 5.

[12]) Vgl. z. B. die Arbeiten von R. Remmert [17] und K. Stein [18].

Definition 5. *Es sei $\varphi: X \to Y$ eine stetige Abbildung eines topologischen Raumes X in einen topologischen Raum Y. (Globaler) Schnitt (bezüglich φ) heiße dann jede stetige Abbildung $\sigma: Y \to X$, für welche $\varphi \circ \sigma$ die identische Abbildung von Y auf sich ist. — Wir werden sagen, die Abbildung φ besitzt einen lokalen Schnitt, wenn es zu jedem Punkt $y \in \varphi(X)$ eine Umgebung V von y in Y und eine stetige Abbildung $\sigma_V: V \to X$ gibt, für welche $\varphi \circ \sigma_V$ die identische Abbildung von V auf sich ist.*

Aus der Existenz eines Schnittes σ folgt, daß die Abbildung φ den Raum X auf Y abbildet. Der Unterraum $S = \sigma(Y)$ ist dann zu Y homöomorph (vermöge der Restriktion von φ auf S). Ferner ist S in X abgeschlossen, falls der Raum X separiert ist. Es sei nämlich $x \in \bar{S}$ ein Punkt der abgeschlossenen Hülle von S in X, $\mathfrak{V}$ der Umgebungsfilter von x in X und $\mathfrak{F}$ die Spur von $\mathfrak{V}$ in S. Da $\mathfrak{F}$ gegen x konvergiert und die Abbildungen φ und σ stetig sind, konvergiert $\varphi(\mathfrak{F})$ gegen $\varphi(x)$ und $\mathfrak{F} = \sigma(\varphi(\mathfrak{F}))$ gegen $\sigma(\varphi(x))$. Wegen der Separiertheit von X ist der Limes von $\mathfrak{F}$ eindeutig bestimmt, also $x = \sigma(\varphi(x))$, also $x \in S$ und somit $S = \bar{S}$.

Aus der Existenz eines lokalen Schnittes folgt die Offenheit von $\varphi(X)$ in Y. Wird also Y zusätzlich als lokal-kompakt vorausgesetzt, so ist auch $\varphi(X)$ lokal-kompakt.

Die Bedeutung dieser Begriffsbildungen für unser Problem erhellt aus dem

Satz 7. *Es sei $\varphi: X \to Y$ eine stetige Abbildung eines lokal-kompakten Raumes X in einen lokal-kompakten Raum Y. Besitzt dann φ einen lokalen Schnitt, so ist φ konservativ.*

Beweis. Zu jedem Punkt $y \in \varphi(X)$ gibt es nach Voraussetzung eine Umgebung V von y in Y und eine stetige Abbildung $\sigma_V: V \to X$, für welche $\varphi \circ \sigma_V$ die identische Abbildung von V auf sich ist. Die Abbildung $\varphi'_V: U \to V$, die auf $U = \overset{-1}{\varphi}(V)$ mit φ übereinstimmt, besitzt dann einen globalen Schnitt. Es ist nämlich $\sigma_V(V) \subset U$ und somit diejenige Abbildung $\sigma'_V: V \to U$, die auf V mit σ_V übereinstimmt, ein Schnitt bezüglich φ'_V. Da V ohne Beschränkung der Allgemeinheit offen gewählt werden kann, lehrt das Lokalisationsprinzip (Satz 1), daß es genügt, unseren Satz unter der Voraussetzung zu beweisen, daß ein Schnitt $\sigma: Y \to X$ bezüglich φ existiert. Dies werde von jetzt an vorausgesetzt.

Wie bereits im Anschluß an die Definition 5 bemerkt wurde, ist dann $S = \sigma(Y)$ in X abgeschlossen; ferner ist die Restriktion von φ auf S eine Homöomorphie von S auf Y. Für jede Menge $K \subset X$ ist aber $\overset{-1}{\sigma}(K) = \varphi(S \cap K)$ und somit kompakt in Y, falls K kompakt in X ist. Es ist daher σ eine eigentliche Abbildung und deswegen $\mu = \sigma(\nu)$ für jedes Maß $\nu \in \mathcal{M}(Y) = \mathcal{M}(Y, \varphi(X))$ definiert. Da $\varphi \circ \sigma$ die identische Abbildung von Y auf sich ist, folgt aus dem Transitivitätsgesetz für Bildmaße, daß mit $\mu = \sigma(\nu)$ und $\nu = (\varphi \circ \sigma)(\nu)$ auch $\varphi(\mu)$ definiert und gleich ν ist. Es ist daher jedes Maß aus $\mathcal{M}(Y)$ φ-darstellbar, also φ konservativ.

Beispiel. Es sei X ein lokal-kompakter, *allgemeiner Faserraum* im Sinne von H. CARTAN [10] (vgl. insbesondere Exposé 6), Y die dann notwendig lokal-kompakte Basis von X, $p: X \to Y$ die Projektionsabbildung und F die Faser. Ist dann X lokal-trivial im Sinne von CARTAN (loc. cit.), d. h. ist X in geeigneter Weise im Kleinen zu einem Produktraum $V \times F$ homöomorph, so besitzt p einen lokalen Schnitt. Nach Satz 7 ist dann p eine konservative Abbildung.

§ 6. Faktorisierung konservativer Abbildungen

6.1. *Ein Transitivitätsgesetz.* — Für die folgenden Betrachtungen wird es sich als zweckmäßig erweisen, zunächst zu untersuchen, ob eine durch Zusammensetzen konservativer Abbildungen entstehende Abbildung wieder konservativ ist.

Dies ist ohne Zusatzvoraussetzung nicht der Fall, denn es kann *jede* stetige Abbildung $\varphi: X \to Y$ eines lokal-kompakten Raumes X in einen lokal-kompakten Raum Y durch Zusammensetzen zweier konservativer stetiger Abbildungen gewonnen werden. In der Tat: es sei $\tilde{X}$ der zur Abbildung φ gehörige lokal-kompakte Raum des Satzes 3, $j: X \to \tilde{X}$ die kanonische Injektion von X in $\tilde{X}$ und $\tilde{\varphi}: \tilde{X} \to Y$ die stetige Fortsetzung von φ auf $\tilde{X}$. Nach Nr. 1.6 ist die Abbildung j konservativ, da X in $\tilde{X}$ offen, also ein lokal-kompakter Unterraum von $\tilde{X}$ ist. Nach Satz 3 ist die Abbildung $\tilde{\varphi}$ eigentlich und als solche nach Satz 2 ebenfalls konservativ. Da $\tilde{\varphi}$ eine Fortsetzung von φ ist, gilt aber $\varphi = \tilde{\varphi} \circ j$, d. h. die beliebige stetige Abbildung φ läßt sich aus zwei konservativen, stetigen Abbildungen zusammensetzen. Besonders hervorgehoben sei die Tatsache, daß $j(X) = X$ in $\tilde{X}$ absolut meßbar, d. h. meßbar in bezug auf jedes positive Maß auf $\tilde{X}$ ist.

Immerhin gilt aber folgendes schwächere *Transitivitätsgesetz:*

Satz 8. *Es seien X, Y, Z lokal-kompakte Räume sowie $\varphi: X \to Y$ und $\psi: Y \to Z$ stetige Abbildungen, von denen die erste den Raum X auf den Raum Y abbildet. Sind dann φ und ψ konservativ, so ist auch die zusammengesetzte Abbildung $\psi \circ \varphi$ konservativ. Ist umgekehrt $\psi \circ \varphi$ konservativ, so ist wenigstens ψ konservativ.*

Beweis. Wir setzen zur Abkürzung $\tau = \psi \circ \varphi$ und bemerken, daß $\tau(X) = \psi(Y)$ ist. Dies folgt aus der Voraussetzung: $\varphi(X) = Y$. — Nun seien zunächst φ und ψ konservativ. Dann existiert zu jedem Maß $\varrho \in \mathcal{M}(Z, \tau(X)) = \mathcal{M}(Z, \psi(Y))$ ein Maß $\nu \in \mathcal{M}(Y)$ mit $\psi(\nu) = \varrho$ und zu ν wegen $\varphi(X) = Y$ ein Maß $\mu \in \mathcal{M}(X)$ mit $\varphi(\mu) = \nu$. Nach dem Transitivitätsgesetz der Nr. 1.5 ist dann $\tau(\mu) = \varrho$, also jedes $\varrho \in \mathcal{M}(Z, \tau(X))$ τ-darstellbar; d. h. τ ist konservativ. — Umgekehrt sei jetzt τ konservativ. Nach einem bereits mehrfach wiederholten Schluß (vgl. Beweis von Satz 1) gibt es zu jedem $f \in \mathcal{K}(Y)$ ein $g \in \mathcal{K}_+(Z)$ mit $|f| \leq g \circ \psi$, also mit $|f \circ \varphi| \leq g \circ \tau$. Zu jedem Maß $\varrho \in \mathcal{M}(Z, \tau(X)) = \mathcal{M}(Z, \psi(Y))$ gibt es nach Voraussetzung ein Maß $\mu \in \mathcal{M}(X)$ mit $\tau(\mu) = \varrho$.

Da $f \circ \varphi$ auf X stetig und $g \circ \tau$ μ-integrierbar ist, folgt aus $|f \circ \varphi| \leqq g \circ \tau$, daß auch $f \circ \varphi$ μ-integrierbar ist, und zwar für jedes $f \in \mathcal{K}(Y)$. Also ist neben $\tau(\mu)$ auch $\nu = \varphi(\mu)$ definiert. Nach dem Transitivitätsgesetz von Nr. 1.5 ist dann $\psi(\nu)$ definiert und $\psi(\nu) = \varrho$. Also ist die Abbildung ψ konservativ.

Wir bemerken noch, daß unter den Voraussetzungen des Satzes 8 aus der Konservativität der Abbildung $\psi \circ \varphi$ *nicht* auf die Konservativität von φ geschlossen werden kann. Dies zeigt folgendes Beispiel: Es sei $\varphi : X \to Y$ eine stetige, nicht-konservative Abbildung von X auf Y (etwa die aus Nr. 1.7) und ψ die Abbildung von Y auf den nur aus einem Punkt bestehenden Raum Z. Die Abbildung $\psi \circ \varphi$ bildet X auf Z ab und ist konservativ (etwa weil ein globaler Schnitt existiert).

6.2. *Konservative Basen und Zerlegungen*[13]). — Unter einer *Zerlegung* (oder Klasseneinteilung) Z einer Menge X verstehen wir wie üblich ein System nicht leerer, paarweise fremder Teilmengen von X, die X überdecken. Die Zerlegungen von X entsprechen in bekannter Weise umkehrbar eindeutig den Äquivalenzrelationen in X. Eine Zerlegung Z eines topologischen Raumes X heißt *einfach*, wenn alle Elemente von Z zusammenhängende Teilmengen von X sind.

Für eine stetige Abbildung $\varphi : X \to Y$ eines topologischen Raumes in einen zweiten ist das System der Fasern $\overset{-1}{\varphi}(\varphi(x))$, $x \in X$, dieser Abbildung eine Zerlegung von X, die wir mit $Z(\varphi)$ bezeichnen und welche die *durch φ bestimmte Zerlegung* heißt. Löst man die einzelnen Fasern von φ in deren Zusammenhangskomponenten auf, so entsteht eine einfache Zerlegung $Z'(\varphi)$ von X, welche die *durch φ bestimmte einfache Zerlegung* heißt. Die Elemente von $Z'(\varphi)$ werden die *Niveaumengen* von φ genannt.

Sind X, Y_1 und Y_2 topologische Räume sowie $\varphi_1 : X \to Y_1$ und $\varphi_2 : X \to Y_2$ stetige Abbildungen, so heißt *φ_2 von φ_1 abhängig*, wenn die Zerlegung $Z'(\varphi_1)$ feiner ist als $Z'(\varphi_2)$, d. h. wenn jedes Element von $Z'(\varphi_1)$ in einem Element von $Z'(\varphi_2)$ enthalten ist. Hiermit gleichwertig ist die Aussage, daß φ_2 auf jeder Niveaumenge von φ_1 konstant ist. Wenn sowohl φ_2 von φ_1 als auch φ_1 von φ_2 abhängig, also $Z'(\varphi_1) = Z'(\varphi_2)$ ist, so nennt man die Abbildungen φ_1 und φ_2 *verwandt*.

Die Frage nach allen von einer konservativen, stetigen Abbildung abhängigen, konservativen, stetigen Abbildungen führt zum Begriff der konservativen Basis.

Definition 6. *Es sei $\varphi_0 : X \to Y_0$ eine konservative, stetige Abbildung eines lokal-kompakten Raumes X in einen lokal-kompakten Raum Y_0. Ein Paar (Y^*, Φ) bestehend aus einem lokal-kompakten Raum Y^* und einer stetigen Abbildung $\Phi : X \to Y^*$ heiße eine* **konservative Basis** *von φ_0, wenn es die folgenden Eigenschaften besitzt:*

---

[13]) Vgl. hierzu K. Stein [18], wo analoge Untersuchungen für holomorphe Abbildungen eines komplexen Raumes in einen zweiten angestellt werden. Wir benützen weitgehend die dort verwendete Terminologie.

a) *die Abbildung Φ ist konservativ und mit φ_0 verwandt, ferner ist $\Phi(X) = Y^*$;*

b) *zu jeder von φ_0 abhängigen, konservativen, stetigen Abbildung $\varphi : X \to Y$ von X in einen lokal-kompakten Raum Y gibt es eine konservative, stetige Abbildung $\varphi^* : Y^* \to Y$ derart, daß $\varphi = \varphi^* \circ \Phi$ ist.*

Ist (Y^*, Φ) eine konservative Basis von $\varphi_0 : X \to Y_0$, so ist für jede konservative, stetige Abbildung $\varphi^* : Y^* \to Y$ die Abbildung $\varphi^* \circ \Phi$ von φ_0 abhängig und nach Satz 8 konservativ. Also durchläuft $\varphi = \varphi^* \circ \Phi$ alle konservativen, von φ_0 abhängigen, stetigen Abbildungen von X in lokal-kompakte Räume, wenn φ^* alle konservativen, stetigen Abbildungen von Y^* in lokal-kompakte Räume durchläuft.

Man überlegt sich leicht, daß eine konservative Basis *„bis auf Homöomorphien" eindeutig* bestimmt ist. Genauer heißt dies: Sind (Y_1^*, Φ_1) und (Y_2^*, Φ_2) konservative Basen von φ_0, so gibt es eine Homöomorphie $\psi^* : Y_1^* \to Y_2^*$ von Y_1^* auf Y_2^* mit $\Phi_2 = \psi^* \circ \Phi_1$. Ferner ist mit (Y_1^*, Φ_1) stets auch $(Y_2^*, \psi^* \circ \Phi_1)$ eine konservative Basis von φ_0, sofern $\psi^* : Y_1^* \to Y_2^*$ eine Homöomorphie von Y_1^* auf einen Raum Y_2^* ist.

Um die Existenz konservativer Basen beweisen zu können, benötigen wir noch folgenden Begriff:

Definition 7. *Eine* **Zerlegung** *Z eines lokal-kompakten Raumes X heiße* **konservativ**, *wenn der Quotientenraum X/Z lokal-kompakt und die kanonische Abbildung $\Phi : X \to X/Z$ konservativ ist.*

Wir erinnern noch daran, daß gemäß der Definition des Quotientenraumes die Abbildung Φ stetig und $Z(\Phi) = Z$ ist.

Beispiele konservativer Zerlegungen liefern die eigentlichen Abbildungen. Es sei nämlich $\varphi : X \to Y$ eine eigentliche Abbildung eines lokal-kompakten Raumes X in einen lokal-kompakten Raum Y. Dann ist die durch φ bestimmte Zerlegung $Z(\varphi)$ von X konservativ. Nach N. Bourbaki [7], p. 104, prop. 17, ist nämlich der Raum $X/Z(\varphi)$ lokal-kompakt und die kanonische Abbildung $\Phi : X \to X/Z(\varphi)$ eigentlich, nach Satz 2 also die Zerlegung $Z(\varphi)$ konservativ.

Die Frage nach der Existenz konservativer Basen beleuchtet der

Hilfssatz 5. *Es sei $\varphi_0 : X \to Y_0$ eine konservative, stetige Abbildung eines lokal-kompakten Raumes X in einen lokal-kompakten Raum Y_0, deren zugehörige einfache Zerlegung $Z'(\varphi_0)$ konservativ ist. Bezeichnet dann Y^* den Quotientenraum $X/Z'(\varphi_0)$ und $\Phi : X \to Y^*$ die kanonische Abbildung, so ist (Y^*, Φ) eine konservative Basis von φ_0.*

Beweis. Nach Voraussetzung ist $Z'(\varphi_0)$ und damit Φ konservativ; ferner ist $\Phi(X) = Y^*$ und Φ mit φ_0 verwandt. Jede stetige, von φ_0 abhängige Abbildung $\varphi : X \to Y$ von X in einen lokal-kompakten Raum Y besitzt genau eine Faktorisierung $\varphi = \varphi^* \circ \Phi$ mit stetigem $\varphi^* : Y^* \to Y$. Da nach Satz 8

mit φ auch φ^* konservativ ist, so ist also (Y^*, Φ) wie behauptet eine konservative Basis.

Nun hat K. Stein [18] (Sätze 5 und 6) folgenden Satz bewiesen: *Es sei $\varphi: X \to Y$ eine stetige Abbildung eines lokal-kompakten Raumes X in einen lokal-kompakten Raum Y, deren sämtliche Niveaumengen kompakt sind. Dann ist die durch φ bestimmte einfache Zerlegung $Z'(\varphi)$ eigentlich, d. h. es ist $X/Z'(\varphi)$ lokal-kompakt und die kanonische Abbildung $\Phi: X \to X/Z'(\varphi)$ eigentlich.*

Hieraus, aus Satz 2 und dem Hilfssatz 5 folgt dann sofort

Satz 9. *Sind die Niveaumengen einer konservativen, stetigen Abbildung $\varphi_0: X \to Y_0$ eines lokal-kompakten Raumes in einen zweiten sämtlich kompakt, so besitzt φ_0 eine konservative Basis.*

Bemerkung. Genügt die Abbildung $\varphi_0: X \to Y_0$ den Voraussetzungen des Satzes 9 und ist (Y^*, Φ) die zugehörige konservative Basis, so besitzt φ_0 selbst eine Faktorisierung $\varphi_0 = \varphi_0^* \circ \Phi$ in die stetigen, konservativen Abbildungen $\Phi: X \to Y^*$ und $\varphi_0^*: Y^* \to Y_0$. Der in [4] verallgemeinerte Faktorisierungssatz von G. T. Whyburn [19], p. 141, kennzeichnet diese Faktorisierung durch topologische Eigenschaften der Faktoren Φ und φ_0^*. Es erweist sich nämlich Φ bzw. φ_0^* als *monotone* bzw. *lichte* Abbildung, d. h. die Fasern der Abbildung sind zusammenhängend bzw. total-unzusammenhängend; außerdem ist Φ eigentlich und eine Abbildung *auf* Y^*. Durch die genannten Eigenschaften der Faktoren sind diese und der Faktorraum Y^* „bis auf Homöomorphien" eindeutig bestimmt.

6.3. *Bemerkung zur Theorie des Lebesgueschen Flächenmaßes.* — Nach einer mündlichen Mitteilung von Herrn W. Fleming gestattet der Satz 2 dieser Arbeit die folgende Anwendung: Es sei X eine im Sinne von L. Cesari [11] zulässige, kompakte Teilmenge der Ebene $\mathbf{R}^2$, also z. B. das von einer einfach geschlossenen Jordan-Kurve umschlossene Gebiet einschließlich seiner Randkurve; ferner sei $\varphi: X \to \mathbf{R}^3$ eine stetige Abbildung von X in den $\mathbf{R}^3$. Die zugehörige parametrische Fläche (φ, X) besitze endliches Lebesguesches Flächenmaß. Bezeichnet $\varphi = \varphi^* \circ \Phi$ die monoton-lichte Whyburnsche Faktorisierung und $Y^* = X/Z'(\varphi)$ den (hier kompakten, metrisierbaren) Whyburnschen Faktorisierungsraum, so existiert nach A. Rosenthal (vgl. [11], p. 403) genau ein positives (Radonsches) Maß μ^* auf Y^* derart, daß für jede offene Menge G^* in Y^* gilt: $\mu^*(G^*)$ ist das Lebesguesche Flächenmaß der Teilfläche $(\varphi', \Phi^{-1}(G^*))$ von (φ, X), wenn hierbei φ' die Restriktion von φ auf $\Phi^{-1}(G^*)$ bedeutet. In diesem Sinne wird bei L. Cesari [11] das Lebesguesche Flächenmaß als Radonsches Maß gedeutet. Da die Abbildung Φ eigentlich ist, folgt aber aus Satz 2, daß auf dem Raum X selbst mindestens ein positives Radonsches Maß μ existiert mit $\Phi(\mu) = \mu^*$. Dann aber gilt für jede bezüglich $Z'(\varphi)$ gesättigte offene Menge G in X, also jede offene Menge $G \subset X$ mit $\Phi^{-1}(\Phi(G)) = G$: es ist $\mu(G)$ das Lebesguesche Flächenmaß der durch die Restriktion φ_G von φ auf G definierten Teilfläche (φ_G, G) von (φ, X).

Literatur

[1] AUMANN, G.: Über die Erweiterung von additiven monotonen Funktionen auf regulär geordneten Halbgruppen. Arch. der Math. **8**, 422—427 (1957). — [2] BAUER, H.: Sur l'existence de mesures avec une image donnée. C. R. Acad. Sci. (Paris) **246**, 1953—1955 (1958). — [3] BAUER, H.: Über die Fortsetzung positiver Linearformen. Sitz.-Ber. math.-naturw. Kl. Bayer. Akad. Wiss. München **1957**, 177—190 (1958). — [4] BAUER, H.: Verallgemeinerung eines Faktorisierungssatzes von G. T. WHYBURN. Arch. der Math. **10**, 373—378 (1959). — [5] BOURBAKI, N.: Intégration, Chap. I—IV. Actual. Scient. et Ind. **1175**, Paris 1952. — [6] BOURBAKI, N.: Intégration, Chap. V. Actual. Sci. et Ind. **1244**, Paris 1956. — [7] BOURBAKI, N.: Topologie générale, Chap. I—II. Actual. Sci. et Ind. **1142**, Paris 1951. — [8] BOURBAKI, N.: Topologie générale, Chap. III—IV. Actual. Sci. et Ind. **1143**, Paris 1951. — [9] BOURBAKI, N.: Espaces vectoriels topologiques, Chap. I—II. Actual. Sci. et Ind. **1189**, Paris 1953. — [10] CARTAN, H.: Espaces fibrés et homotopie. Séminaire E. N. S. **1949/50**, 2e édition. — [11] CESARI, L.: Surface Area. Ann. of Math. Studies **35**, Princeton 1956. — [12] MORSE, M., and W. TRANSUE: Semi-normed vector spaces with duals of integral type. J. d'Analyse Math. **4**, 149—186 (1954/55). — [13] NAKANO, H.: On the product of relative spectra. Ann. of Math. **49**, 281—315 (1948). — [14] NAMIOKA, I.: Partially ordered linear topological spaces. Memoirs Amer. math. Soc. **24**, Providence 1957. — [15] NÖBELING, G., u. H. BAUER: Über die Erweiterungen topologischer Räume. Math. Ann. **130**, 20—45 (1955). — [16] NÖBELING, G., u. H. BAUER: Allgemeine Approximationskriterien mit Anwendungen. J.-Ber. Dtsch. Math.-Verein. **58**, 54—72 (1955). — [17] REMMERT, R.: Holomorphe und meromorphe Abbildungen komplexer Räume. Math. Ann. **133**, 328—374 (1957). — [18] STEIN, K.: Analytische Zerlegungen komplexer Räume. Math. Ann. **132**, 63—93 (1956). — [19] WHYBURN, G. T.: Analytic Topology. Amer. math. Soc. Colloq. Publ. **28**, New York 1942.

(Eingegangen am 1. Juni 1959)

HOLMANN, H.
Math. Annalen 138, 428—441 (1959)

Zur Abbildungstheorie komplexer Mannigfaltigkeiten

Von
HARALD HOLMANN in Münster (Westf.)

Einleitung

In der Funktionentheorie mehrerer komplexer Veränderlicher wird die Rolle des Kreises der klassischen Theorie weitgehend von den sogenannten kreissymmetrischen Gebieten [2] übernommen, das sind solche Gebiete des C^n, die gegenüber gewissen kompakten Lieschen Gruppen linearer Transformationen, die den Nullpunkt festlassen, invariant bleiben. Aus diesem Grunde ist man in der Abbildungstheorie an der Frage interessiert, wann ein Gebiet des C^n sich auf ein kreissymmetrisches eineindeutig und holomorph abbilden läßt. Da es jedoch Gebiete (sogar einfach zusammenhängende Holomorphiegebiete) im C^n gibt, die vollkommen starr sind, d. h. keine holomorphen Automorphismen außer der Identität besitzen, muß man die Untersuchung von vorneherein auf solche Gebiete beschränken, die eine unendliche kompakte Liesche Gruppe holomorpher Automorphismen zulassen. Für beschränkte Gebiete des C^2 wurde das Problem 1931 von H. CARTAN [3] gelöst. Er bewies den folgenden nach ihm benannten Abbildungssatz: Jedes beschränkte Gebiet des C^2, das eine kompakte Liesche Automorphismengruppe (mit unendlich vielen Elementen) mit einem Fixpunkt zuläßt, ist zu einem (m, p)-Gebiet (auch Cartanscher Körper genannt) holomorph äquivalent. Unter einem (m, p)-Gebiet wird dabei ein spezielles kreissymmetrisches Gebiet verstanden, das gegenüber den linearen Transformationen $z_1 \to z_1 e^{i m \theta}$, $z_2 \to z_2 e^{i p \theta}$, $(m, p) = 1$, $0 \leq \theta \leq 2\pi$, invariant bleibt. Beim Beweis wird ganz entscheidend davon Gebrauch gemacht, daß die Gebiete 2-dimensional sind und Fixpunkte bezüglich der Automorphismengruppe besitzen.

In der vorliegenden Arbeit soll mit anderen Mitteln eine Verallgemeinerung des Cartanschen Abbildungssatzes für n-dimensionale komplexe Mannigfaltigkeiten gegeben werden, wobei wir nur voraussetzen wollen, daß die Mannigfaltigkeiten unendliche kompakte Liesche Gruppen holomorpher Automorphismen zulassen (die Forderung der Existenz von Fixpunkten wird ebenfalls fortgelassen). Die Vermutung, daß man hierzu nur die (m, p)-Gebiete durch (s. Definition 2) $(p_1, \ldots, p_n)$-Gebiete zu ersetzen hat, die analog zu den (m, p)-Gebieten definiert werden können, erweist sich als falsch. Selbst unter der Voraussetzung, daß die Mannigfaltigkeiten holomorph-vollständig sind und Fixpunkte bezüglich der Automorphismengruppe besitzen, lassen sich Gegenbeispiele zu einer so einfachen Verallgemeinerung des Abbildungssatzes konstruieren. Um zum Ziele zu kommen, wird man also eine Ver-

allgemeinerung der $(p_1, \ldots, p_n)$-Gebiete vornehmen müssen. Es erweist sich als zweckmäßig, sogenannte Cartansche Mannigfaltigkeiten zu betrachten, das sind solche komplexe Mannigfaltigkeiten, die einen komplexen Atlas besitzen, der ihnen lokal die Struktur eines Cartanschen Körpers aufprägt (s. Definition 3). Wir können dann eine Verallgemeinerung des Cartanschen Abbildungssatzes in folgender Form aussprechen (s. Satz 2):
Jede komplexe Mannigfaltigkeit, die eine unendliche kompakte Liesche Gruppe holomorpher Automorphismen besitzt, ist holomorph äquivalent zu einer Cartanschen Mannigfaltigkeit.

§ 1. Cartansche Mannigfaltigkeiten

Es sollen zunächst einige Begriffe zusammengestellt werden. Unter einer n-dimensionalen komplexen Mannigfaltigkeit X^n verstehen wir einen Hausdorffschen Raum mit einem n-dimensionalen komplexen Atlas. Dabei ist ein n-dimensionaler komplexer Atlas von X^n definiert als eine Kollektion $\{(X_i, \varphi_i), i \in I\}$, wo $\{X_i, i \in I\}$ eine offene Überdeckung von X^n ist und φ_i eine topologische Abbildung von X_i auf eine offene Menge des n-dimensionalen komplexen Vektorraumes C^n darstellt, derart daß die Abbildung

$$\varphi_{ij} := \varphi_i \circ \varphi_j^{-1} \mid \varphi_j(X_i \cap X_j) \to \varphi_i(X_i \cap X_j)$$

holomorph ist. (X_i, φ_i) wird dabei eine n-dimensionale komplexe Karte von X_i genannt.

Eine q-dimensionale Liesche Gruppe ist definiert als eine topologische Gruppe L^q, die einen q-dimensionalen reell-analytischen Atlas besitzt, derart daß die Abbildung $(l_1, l_2) \to l_1 \circ l_2^{-1}$ von $L^q \times L^q$ auf L^q reell-analytisch ist. In Analogie zur Definition eines komplexen Atlanten ist ein q-dimensionaler reell-analytischer Atlas von L^q erklärt als eine Kollektion von Karten $\{(L_i, \psi_i), i \in I\}$, wobei $\{L_i, i \in I\}$ eine offene Überdeckung von L^q darstellt und ψ_i die offene Menge L_i topologisch auf eine offene Menge des q-dimensionalen reellen Vektorraumes E^q abbildet, so daß die Abbildung

$$\psi_{ij} := \psi_i \circ \psi_j^{-1} \mid \psi_j(L_i \cap L_j) \to \psi_i(L_i \cap L_j)$$

reell-analytisch ist.

Die holomorphen Automorphismen einer komplexen Mannigfaltigkeit X^n bilden eine Gruppe G, die durch die Topologie der „gleichmäßigen Konvergenz im Inneren" zu einer topologischen Gruppe gemacht wird. Eine Untergruppe L von G wird Liesche Automorphismengruppe genannt, wenn sie (mit dieser Topologie versehen) einen reell-analytischen Atlas besitzt, der sie zu einer Lieschen Gruppe macht, und wenn die Abbildung $(l, x) \to l(x)$ von $L \times X^n$ auf X^n reell-analytisch ist. Da L eine reell-analytische Mannigfaltigkeit ist und die n-dimensionale komplexe Mannigfaltigkeit X^n als eine $2n$-dimensionale reell-analytische Mannigfaltigkeit interpretiert werden kann, ist es sinnvoll, von einer reell-analytischen Abbildung von $L \times X^n$ auf X^n zu sprechen.

Wir stellen uns die Aufgabe, für komplexe Mannigfaltigkeiten, die eine kompakte Liesche Automorphismengruppe (mit unendlich vielen Elementen)

besitzen, eine Verallgemeinerung des Cartanschen Abbildungssatzes zu geben. Man kann sich dabei auf die Untersuchung von k_1-Mannigfaltigkeiten beschränken.

Definition 1: Unter einer k_1-Mannigfaltigkeit verstehen wir eine komplexe Mannigfaltigkeit X^n mit einer 1-dimensionalen kompakten Lieschen Automorphismengruppe L^1.

Jede kompakte Liesche Gruppe (mit unendlich vielen Elementen) enthält nämlich mindestens eine globale 1-dimensionale kompakte Liesche Untergruppe L^1. Wir können dabei annehmen, daß die Transformationen von L^1 in kanonischer Form

$$(1) \qquad P \to P^* = F_\theta(P), \; P, P^* \in X^n, \; 0 \leqq \theta \leqq 2\pi \,,$$

gegeben sind, d. h. es gilt:

$$(2) \qquad F_{\theta^*} \circ F_\theta(P) = F_{\theta^* + \theta}(P) \,,$$

wobei die Summe $\theta^* + \theta$ modulo 2π gerechnet wird. Setzen wir $\theta = -i \log y$, so lassen sich die Transformationen (1) auch in der Form:

$$(3) \qquad P \to P^* = H_y(P) := F_{-i \log y}(P), \; |y| = 1 \,,$$

schreiben. Gleichung (2) geht dabei über in

$$(4) \qquad H_{y^*} \circ H_y(P) = H_{y^* y}(P) \,.$$

Bei einer Ausdehnung des Cartanschen Abbildungssatzes auf k_1-Mannigfaltigkeiten muß man sich überlegen, was den Cartanschen Körpern hier entspricht. In Verallgemeinerung der Cartanschen Körper werden wir sogenannte Cartansche Mannigfaltigkeiten definieren. Zuvor müssen wir jedoch erklären, was wir unter einem $(p_1, \ldots, p_n)$-Gebiet verstehen.

Definition 2: Ein Gebiet U des C^n heißt ein $(p_1, \ldots, p_n)$-Gebiet, wenn es gegenüber den linearen Transformationen $A_y(p_1, \ldots, p_n)$:

$$(5) \qquad z_\nu \to z_\nu^* = z_\nu y^{p_\nu}, \quad |y| = 1 \,, \quad \nu = 1, \ldots, n \,,$$

invariant bleibt. Dabei sind $p_1, \ldots, p_n$ ganze Zahlen, deren g. g. T. gleich eins ist.

Es sei angemerkt, daß nicht gefordert wird, daß der Nullpunkt im Inneren des Gebietes U liegt.

Definition 3: Unter einer Cartanschen Mannigfaltigkeit verstehen wir eine komplexe Mannigfaltigkeit X^n mit einem komplexen Atlas $\{(X_i, \varphi_i), i \in I\}$, der die folgenden Eigenschaften besitzt:

1) $\varphi_i(X_i)$ und $\varphi_i(X_i \cap X_j)$ sind $(p_1^i, \ldots, p_n^i)$-Gebiete im C^n.

2) Die Abbildungen $\varphi_{ij} := \varphi_i \circ \varphi_j^{-1}$, $i, j \in I$, sind mit den Automorphismen ${}^iA_y := A_y(p_1^i, \ldots, p_n^i)$ der $(p_1^i, \ldots, p_n^i)$-Gebiete $\varphi_i(X_i)$, $i \in I$, verträglich, d. h. es gilt:

$$(6) \qquad {}^iA_y \circ \varphi_{ij} = \varphi_{ij} \circ {}^jA_y \,.$$

Die Gleichung (6) von Definition 3 ist äquivalent zu: $\varphi_i^{-1} \circ {}^iA_y \circ \varphi_i = \varphi_j^{-1} \circ {}^jA_y \circ \varphi_j$. Somit stellen die Abbildungen $P \to P^* = A_y(P)$, $P, P^* \in X^n$,

$|y| = 1$, holomorphe Automorphismen von X^n dar, wenn wir $A_y(P)$ für $P \in X_i$ jeweils durch $A_y(P) := \varphi_i^{-1} \circ {}^iA_y \circ \varphi_i(P)$ definieren. Die Menge aller Automorphismen A_y, $|y| = 1$, bildet eine 1-dimensionale kompakte Liesche Automorphismengruppe $\Lambda(X^n)$, von der wir sagen, daß sie X^n in kanonischer Weise zugeordnet ist. Die Cartanschen Mannigfaltigkeiten stellen somit eine Klasse von k_1-Mannigfaltigkeiten dar.

Den Hauptsatz der vorliegenden Arbeit können wir jetzt folgendermaßen formulieren:

Satz 1: *Jede k_1-Mannigfaltigkeit ist zu einer Cartanschen Mannigfaltigkeit holomorph äquivalent.*

Da jede komplexe Mannigfaltigkeit, die eine kompakte Liesche Automorphismengruppe (mit unendlich vielen Elementen) besitzt, eine k_1-Mannigfaltigkeit ist, so läßt sich Satz 1 auch folgendermaßen schreiben:

Satz 2: *Jede komplexe Mannigfaltigkeit, die eine kompakte Liesche Automorphismengruppe (mit unendlich vielen Elementen) besitzt, ist holomorph äquivalent zu einer Cartanschen Mannigfaltigkeit.*

Beispiele: Eine Cartansche Mannigfaltigkeit, die zu einem $(p_1, \ldots, p_n)$-Gebiet holomorph äquivalent ist, wollen wir als trivial bezeichnen. Der n-dimensionale komplexe projektive Raum P^n (mit den homogenen Koordinaten $[w] = [w_0, \ldots, w_n]$) stellt ein einfaches Beispiel einer nicht trivialen Cartanschen Mannigfaltigkeit dar. Die Transformationen der kanonisch zugeordneten Automorphismengruppe $\Lambda(P^n)$ können wie folgt gewählt werden: $A_y : [w_0, \ldots, w_n] \to [w_0 y^{p_0}, \ldots, w_n y^{p_n}]$, wobei $|y| = 1$ und der g. g. T. von $p_1 - p_0, \ldots, p_n - p_0$ gleich eins ist. Einen Atlas $\{(X_\nu, \varphi_\nu), \nu = 0, \ldots, n\}$ von P^n, der die Bedingungen 1) und 2) von Definition 3 erfüllt, erhält man folgendermaßen. Wir definieren $X_\nu := \{[w] : [w] \in P^n, w_\nu \neq 0\}$. Die Abbildung $\varphi_\nu | X_\nu \to C^n$ sei gegeben durch

$$\varphi_\nu : [w] \to {}^\nu z := ({}^\nu z_1, \ldots, {}^\nu z_n) := \left(\frac{w_0}{w_\nu}, \ldots, \frac{w_{\nu-1}}{w_\nu}, \frac{w_{\nu+1}}{w_\nu}, \ldots, \frac{w_n}{w_\nu}\right).$$

Man rechnet leicht aus, daß die Abbildungen ${}^\nu A_y := \varphi_\nu \circ A_y \circ \varphi_\nu^{-1}$, $|y| = 1$, holomorphe Automorphismen von $\varphi_\nu(X_\nu)$ und $\varphi_\nu(X_\nu \cap X_\mu)$ darstellen. Es gilt:

$${}^\nu A_y : {}^\nu z \to ({}^\nu z_1 y^{p_0 - p_\nu}, \ldots, {}^\nu z_\nu y^{p_{\nu-1} - p_\nu}, {}^\nu z_{\nu+1} y^{p_{\nu+1} - p_\nu}, \ldots, {}^\nu z_n y^{p_n - p_\nu}),$$

d. h. $\varphi_\nu(X_\nu)$ und $\varphi_\nu(X_\nu \cap X_\mu)$ sind $(p_0 - p_\nu, \ldots, p_{\nu-1} - p_\nu, p_{\nu+1} - p_\nu, \ldots, p_n - p_\nu)$-Gebiete.

Setzt man $\varphi_{\nu\mu} := \varphi_\nu \circ \varphi_\mu^{-1}$, so ergibt sich sofort folgende Gleichung: $\varphi_{\nu\mu} \circ {}^\mu A_y = {}^\nu A_y \circ \varphi_{\nu\mu}$, d. h. Bedingung 2) von Definition 3 ist erfüllt, und P^n mit dem angegebenen Atlas stellt somit eine Cartansche Mannigfaltigkeit dar.

Ein zweites Beispiel einer Cartanschen Mannigfaltigkeit erhalten wir durch eine Modifikation des C^n in der Ebene $\{z = (z_1, \ldots, z_n) : z_1 = \cdots = z_{k+1} = 0\}$. Wir betrachten den Graphen G_ψ ([5], Seite 367) einer meromorphen Abbildung $\psi | C^n \to P^k$, $k < n$, die folgendermaßen gegeben sei: $\psi : z \to [w] := [z_1, \ldots, z_{k+1}]$, falls nicht $z_1 = \cdots = z_{k+1} = 0$, $\psi : (0, \ldots, 0, z_{k+2}, \ldots, z_n) \to \{[w] : [w] \in P^k\}$. Dabei sind die homogenen Koordinaten von P^k mit $[w] = [w_1, \ldots, w_{k+1}]$

bezeichnet. G_ν stellt eine n-dimensionale komplexe Mannigfaltigkeit dar. Die G_ν kanonisch zugeordnete Automorphismengruppe $A(G_\nu)$, die G_ν zu einer Cartanschen Mannigfaltigkeit macht, erhalten wir auf folgende Weise. Wir definieren zunächst Automorphismen A_ν des C^n:

$$A_\nu : z \to (z_1 y^{p_1}, \ldots, z_n y^{p_n}), \; |y| = 1 ,$$

und Automorphismen B_ν von P^k:

$$B_\nu : [w] \to [w_1 y^{p_1}, \ldots, w_{k+1} y^{p_{k+1}}], \; |y| = 1 .$$

Dabei sind $p_1, \ldots, p_n$ ganze Zahlen, deren g. g. T. gleich eins ist. $A_\nu \times B_\nu$ stellt dann einen holomorphen Automorphismus von $C^n \times P^k$ dar. $A_\nu \times B_\nu$ ist definiert durch:

$$A_\nu \times B_\nu : z \times [w] \to A_\nu(z) \times B_\nu([w]) .$$

Wir behaupten nun, daß der Graph G_ν durch $A_\nu \times B_\nu$ eineindeutig und holomorph auf sich abgebildet wird. Zu diesem Zweck haben wir nur zu zeigen, daß $A_\nu \times B_\nu(z \times \psi(z)) \subset G_\nu$ für alle $z \in C^n$. Man rechnet sofort aus, daß gilt:

$$\psi \circ A_\nu = B_\nu \circ \psi .$$

Daraus folgt dann: $A_\nu \times B_\nu : z \times \psi(z) \to A_\nu(z) \times B_\nu \circ \psi(z) = A_\nu(z) \times \psi \circ A_\nu(z) = z' \times \psi(z') \subset G$, q. e. d.

Eine weitere Klasse von Cartanschen Mannigfaltigkeiten stellen die komplexen Vektorraumbündel über einer komplexen Mannigfaltigkeit dar.

V^k sei ein k-dimensionales Vektorraumbündel über der komplexen Mannigfaltigkeit X^n, d. h. V^k und X^n besitzen komplexe Atlanten $\{(V_i, \psi_i), i \in I\}$ bzw. $\{(X_i, \varphi_i), i \in I\}$ mit folgenden Eigenschaften:

1) ψ_i bildet V_i bzw. $V_i \cap V_j$ eineindeutig und holomorph auf $\varphi_i(X_i) \times C^k$ bzw. $\varphi_i(X_i \cap X_j) \times C^k$ ab.

2) Die Abbildung $\psi_{ij} := \psi_i \circ \psi_j^{-1} | \varphi_j(X_i \cap X_j) \times C^k \to \varphi_i(X_i \cap X_j) \times C^k$ hat die Form:

$$\psi_{ij} : {}^j x \times {}^j z \to {}^i x \times {}^i z = \varphi_{ij}({}^j x) \times C_{ij}({}^j x) \circ {}^j z .$$

Dabei ist C_{ij} eine holomorphe Abbildung von $\varphi_j(X_i \cap X_j)$ in $GL(k, C)$.

$\psi_i(V_i) = \varphi_i(X_i) \times C^k$ läßt die folgenden holomorphen Automorphismen A_ν, $|y| = 1$, zu:

$$A_\nu : {}^i x \times {}^i z \to {}^i x' \times {}^i z' = {}^i x \times y\, {}^i z .$$

Hierdurch werden $\psi_i(V_i)$ und $\psi_i(V_i \cap V_j)$ zu $(p_1, \ldots, p_{n+k})$-Gebieten im C^{n+k}, wobei gilt:

$$p_1 = \cdots = p_n = 0 \text{ und } p_{n+1} = \cdots = p_{n+k} = 1 .$$

Da auch noch gilt $A_\nu \circ \psi_{ij} = \psi_{ij} \circ A_\nu$, so besitzt der Atlas $\{(V_i, \psi_i), i \in I\}$ von V^k die Eigenschaften 1) und 2) von Definition 3. V^k ist somit eine Cartansche Mannigfaltigkeit.

§ 2. Kompakte 1-dimensionale Liesche Automorphismengruppen

X^n sei eine k_1-Mannigfaltigkeit, d. h. sie besitzt eine 1-dimensionale kompakte Liesche Automorphismengruppe L^1, deren Elemente in der Form (3) vorliegen. Bevor wir Satz 1 beweisen, sollen einige Untersuchungen über die

Gruppe L^1 und ihre zyklischen Untergruppen endlicher Ordnung angestellt werden. Wir wollen zuerst das folgende Lemma beweisen, das Auskunft gibt über eine holomorphe Fortsetzung der Abbildung:

$$H : (P, y) \to P^* = H_y(P) \tag{7}$$

von $X^n \times \{y : |y| = 1\}$ auf X^n.

Lemma 1: *Für jedes Gebiet $X \subset X^n$ gibt es ein $\varepsilon > 0$, so daß Abbildung* (7) *eindeutig fortgesetzt werden kann zu einer holomorphen Abbildung von $X \times \{y : 1 - \varepsilon < |y| < 1 + \varepsilon\}$ in X^n.*

Beweis: Wir ordnen jedem Punkt $Q \in X^n$ eine offene Umgebung X_Q und eine komplexe Karte (X_Q, φ_Q) so zu, daß $\varphi_Q(Q) = 0$ ist. $\overline{X}$ bezeichne die abgeschlossene Hülle von X, und schließlich sei noch die Abkürzung $W := \overline{X} \times \{y : |y| = 1\}$ eingeführt.

(Q, y^*) sei ein Punkt aus W und $Q^* = H_{y^*}(Q)$ das zugehörige H-Bild in X^n. In den Koordinaten ${}^Qz = ({}^Qz_1, \ldots, {}^Qz_n)$ von $\varphi_Q(X_Q)$ und ${}^{Q^*}z = ({}^{Q^*}z_1, \ldots, {}^{Q^*}z_n)$ von $\varphi_{Q^*}(X_{Q^*})$ läßt sich Abbildung (7) dann folgendermaßen ausdrücken:

$$\mathfrak{H} : ({}^Qz, y) \to {}^{Q^*}z := \mathfrak{H}_y({}^Qz) := \varphi_Q \circ H_y \circ \varphi_Q^{-1}({}^Qz)\,. \tag{8}$$

Die Komponenten $h_\nu({}^Qz, y)$, $\nu = 1, \ldots, n$, von $\mathfrak{H}_y({}^Qz)$ lassen sich in $(n + 1)$-fache Potenzreihen

$$h_\nu({}^Qz, y) = \sum_{\substack{\mu_0 = 0\\ \vdots \\ \mu_n = 0}}^{\infty} {}^Qa_{\mu_0 \ldots \mu_n} (y - y^*)^{\mu_0}\, {}^Qz_1^{\mu_1} \ldots {}^Qz_n^{\mu_n} \tag{9}$$

entwickeln, die in einer hinreichend kleinen Umgebung U_{Q,y^*} von $({}^Qz, y) = (0, y^*)$ konvergieren. Wir können annehmen, daß die Umgebung die Form $U_{Q,y^*} := \{({}^Qz, y) : |{}^Qz| < \varepsilon(Q, y^*),\ |y - y^*| < \varepsilon(Q, y^*)\}$ besitzt, wobei $\varepsilon(Q, y^*)$ eine hinreichend kleine positive Zahl ist. Die Potenzreihenentwicklungen (9) definieren eine eindeutige holomorphe Abbildung $\mathfrak{H}^* \mid U_{Q,y^*} \to C^n$, die folgendermaßen geschrieben werden soll: $\mathfrak{H}^* : ({}^Qz, y) \to {}^{Q^*}z = \mathfrak{H}_y^*({}^Qz)$. Wir wollen annehmen, daß $\varepsilon(Q, y^*)$ so klein gewählt ist, daß $\mathfrak{H}^*(U_{Q,y^*}) \subset \varphi_{Q^*}(X_{Q^*})$. Dann stellt

$$H^* : (P, y) \to P^* := H_y^*(P) := \varphi_Q^{-1} \circ \mathfrak{H}_y^* \circ \varphi_Q(P) \tag{10}$$

eine eindeutige holomorphe Fortsetzung von Abbildung (7) in das Gebiet $W_{Q,y^*} := \{(P, y) : |\varphi_Q(P)| < \varepsilon(Q, y^*),\ |y - y^*| < \varepsilon(Q, y^*)\}$ dar. Die Gebiete W_{Q,y^*}, $(Q, y^*) \in W$ überdecken ganz W. Da W kompakt ist, kann man sogar endlich viele Gebiete W_{Q,y_σ}, $\sigma = 1, \ldots, s$, auswählen, die W überdecken. Es gibt sicherlich ein in $\hat{W} := \bigcup_{\sigma=1}^{s} W_{Q_\sigma, y_\sigma}$ gelegenes Gebiet der Form $X \times \{y : 1 - \varepsilon < |y| < 1 + \varepsilon\}$, wobei ε eine genügend kleine positive Zahl ist. Da sich Abbildung (7) in $\hat{W}$ eindeutig holomorph fortsetzen läßt, so auch in $X \times \{y : 1 - \varepsilon < |y| < 1 + \varepsilon\}$, q. e. d.

In einer k_1-Mannigfaltigkeit X^n liefert die zugehörige kompakte 1-dimensionale Liesche Automorphismengruppe L^1 eine Einteilung der Punkte in verschiedene Klassen.

Definition 4: Ein Punkt P aus der komplexen k_1-Mannigfaltigkeit X^n gehört zur Klasse K_0, wenn P ein Fixpunkt bezüglich der Automorphismengruppe L^1 ist. P gehört zur Klasse K_r, $r > 0$, wenn r die größte positive ganze Zahl ist, so daß $P = H_{\alpha_r}(P)$, wobei $\alpha_r := e^{2\pi i/r}$ ist.

Man überlegt sich sofort, daß P genau dann zur Klasse K_r, $r > 0$, gehört, wenn die Abbildung $C_P: y \to P^* = H_y(P)$, $|y| = 1$, $0 \leqq \arg(y) \leqq 2\pi/r$, eine einfach geschlossene einmal durchlaufene Kurve darstellt. Alle Punkte dieser Kurve gehören dann ebenfalls zur Klasse K_r. Die Menge aller Transformationen

$$
{}^{\varrho}H : P \to P^* := H_{\alpha_r^{\varrho}}(P),\ \varrho = 0, 1, \ldots, r-1\,, \tag{11}
$$

stellt eine zyklische Untergruppe G_r der Ordnung r von L^1 dar, die die Punkte der Klasse K_r fix läßt. Das folgende Lemma liefert eine kanonische Darstellung der Gruppe G_r, von der wir im nächsten Paragraphen beim Beweis von Satz 1 Gebrauch machen werden.

Lemma 2: *X^n sei eine k_1-Mannigfaltigkeit. Jedem Punkt $Q \in X^n$ der Klasse K_r, $r > 0$, läßt sich eine komplexe Karte (X, φ), $Q \in X$, $\varphi(Q) = 0$, mit folgenden Eigenschaften zuordnen:*

1) *X ist invariant gegenüber den Transformationen* (11) *der Gruppe G_r.*

2) *Die Automorphismen ${}^{\varrho}\mathfrak{R} : x \to {}^{\varrho}x := \varphi \circ {}^{\varrho}H \circ \varphi^{-1}(x)$ von $\varphi(X)$ (mit den Koordinaten $x = (x_1, \ldots, x_n)$) haben die Form $x_\nu \to {}^{\varrho}x_\nu = x_\nu \alpha_r^{p_\nu \varrho}$, $\nu = 1, \ldots, n$. Dabei sind die p_ν ganze Zahlen, $0 \leqq p_\nu < r$, und es gilt:* g. g. T. $(r, p_2, \ldots, p_n) = 1$.

3) $\left.\dfrac{\partial k_1(x, y)}{\partial y}\right|_{x=0,\, y=1} \neq 0$, *wobei $k_\nu(x, y)$, $\nu = 1, \ldots, n$, die Komponenten von $\mathfrak{R}_y(x) := \varphi \circ H_y \circ \varphi^{-1}(x)$ darstellen.*

Beweis: (X', φ') sei irgendeine Karte einer Umgebung X' von Q, wobei $\varphi'(X')$ beschränkt und $\varphi'(Q) = 0$ ist. Wegen der Stetigkeit der Abbildungen (11) gibt es für jedes ϱ, $0 \leqq \varrho \leqq r-1$, eine Umgebung ${}^{\varrho}X$ von Q, die durch ${}^{\varrho}H$ in X' abgebildet wird. Setzen wir $X'' := \bigcap_{\varrho=0}^{r-1} {}^{\varrho}X$, so bleibt

$$
\hat{X} := \{P^* : P^* = {}^{\varrho}H(P),\ P \in X'',\ \varrho = 0, \ldots, r-1\} \subset X'
$$

invariant gegenüber den Transformationen der Gruppe G_r. Setzen wir $\hat{\varphi} = \varphi' | \hat{X}$, so stellt $(\hat{X}, \hat{\varphi})$ eine Karte von $\hat{X}$ dar, die der Bedingung 1 von Lemma 2 genügt. In den Koordinaten $z = (z_1, \ldots, z_n)$ von $\hat{\varphi}(\hat{X})$ stellen sich die Abbildungen (11) folgendermaßen dar:

$$
{}^{\varrho}\mathfrak{F} : z \to {}^{\varrho}z := \hat{\varphi} \circ {}^{\varrho}H \circ \hat{\varphi}^{-1}(z)\,. \tag{12}
$$

Die Komponenten ${}^{\varrho}f_\nu(z)$ von ${}^{\varrho}\mathfrak{F}(z)$ lassen sich in einer Umgebung des Nullpunktes in Potenzreihen entwickeln:

$$
{}^{\varrho}f_\nu(z) = \sum_{\mu=1}^{n} {}^{\varrho}a_{\nu\mu} z_\mu + \cdots,\ \nu = 1, \ldots, n\,. \tag{13}
$$

Setzen wir ${}^{\varrho}A = ({}^{\varrho}a_{\nu\mu})$, so läßt sich die Entwicklung (13) auch folgendermaßen schreiben:

$$
{}^{\varrho}\mathfrak{F}(z) = {}^{\varrho}A \circ z + \cdots \tag{14}
$$

Da $\varphi'(X')$ beschränkt ist und wegen $\hat{\varphi}(\hat{X}) \subset \varphi'(X')$ dasselbe für $\hat{\varphi}(\hat{X})$ gilt, so sind auf Grund des Cartanschen Eindeutigkeitssatzes ([1], [3]) die Abbildungen ${}^{\varrho}\mathfrak{F}$ durch ihre Linearteile, d. h. die Matrizen ${}^{\varrho}A$ eindeutig festgelegt. Da bei einer Verknüpfung zweier Abbildungen der Form (14) sich die Linearteile miteinander multiplizieren, so bilden die Matrizen ${}^{\varrho}A$, $\varrho = 0, \ldots, r-1$, eine zu G_r isomorphe zyklische Gruppe Γ_r. Es gilt: ${}^{\varrho}A = {}^{1}A^{\varrho}$. Wir wollen eine nicht-singuläre Matrix R angeben, so daß $R \circ {}^{\varrho}A \circ R^{-1} = {}^{\varrho}D$ Diagonalform erhält. Zu diesem Zweck betrachten wir die positiv definite hermitesche Form:

$$z^* \circ A \circ z, \text{ wobei } A := \sum_{\varrho'=0}^{r-1} {}^{\varrho'}A^* \circ {}^{\varrho'}A . \tag{15}$$

Der Stern bedeutet den Übergang zur transponierten konjugiert komplexen Matrix. Wie man durch Einsetzen sofort sieht, ist die Form (15) gegenüber den linearen Substitutionen $z' = {}^{\varrho}A \circ z$ invariant, d. h. es gilt:

$${}^{\varrho}A^* \circ A \circ {}^{\varrho}A = A . \tag{16}$$

Eine positiv definite hermitesche Matrix läßt sich stets als Quadrat einer anderen positiv definiten hermiteschen Matrix schreiben ([4], Seite 14). Es gibt somit eine positiv definite hermitesche Matrix R_1, so daß $R_1^2 = A$. Man zeigt nun, daß die Matrizen ${}^{\varrho}B := R_1 \circ {}^{\varrho}A \circ R_1^{-1}$ unitär sind, indem man unter Benutzung von Gleichung (16) die Gültigkeit von ${}^{\varrho}B^* \circ {}^{\varrho}B = E$ nachweist. Es gibt dann eine unitäre Transformation R_2, so daß ${}^{1}D := R_2 \circ {}^{1}B \circ R_2^{-1}$ Diagonalgestalt erhält. Setzen wir $R := R_2 \circ R_1$, so können wir auch schreiben: ${}^{1}D = R \circ {}^{1}A \circ R^{-1}$. Für die Matrizen ${}^{\varrho}D := R \circ {}^{\varrho}A \circ R^{-1}$ gilt dann ebenfalls ${}^{\varrho}D = {}^{1}D^{\varrho}$. Sie besitzen also auch Diagonalgestalt. Die Diagonalelemente von ${}^{1}D$ seien mit λ_ν, $\nu = 1, \ldots, n$, bezeichnet. Wegen ${}^{1}D^r = E$ gilt $\lambda_\nu^r = 1$ oder $\lambda_\nu = \alpha_r^{p_\nu}$. $\alpha_r := e^{2\pi i/r}$, wobei die p_ν ganze Zahlen, $0 \leq p_\nu < r$, sind. Die Matrizen ${}^{\varrho}D$ haben somit die Form:

$${}^{\varrho}D = \begin{pmatrix} \alpha_r^{p_1\varrho} & & 0 \\ & \ddots & \\ 0 & & \alpha_r^{p_n\varrho} \end{pmatrix}. \tag{17}$$

Da die ${}^{\varrho}D$, $\varrho = 0, \ldots, r-1$, wieder eine zyklische Gruppe Γ_r' der Ordnung r bilden, so ist der g. g. T. von $r, p_1, \ldots, p_n$ gleich eins. Setzen wir $w = R \circ z$, so geht die Abbildung (12) über in:

$${}^{\varrho}\mathfrak{H} : w \to {}^{\varrho}w := R \circ {}^{\varrho}\mathfrak{F} \circ R^{-1} \circ w = {}^{\varrho}D \circ w + \cdots . \tag{18}$$

Durch die Abbildung

$$\psi : w \to x := \frac{1}{r} \sum_{\varrho=0}^{r-1} {}^{\varrho}D^{-1} \circ {}^{\varrho}\mathfrak{H}(w) = w + \cdots \tag{19}$$

wird eine in $R(\hat{\varphi}(\hat{X}))$ gelegene Umgebung $\hat{V}$ des Nullpunktes eineindeutig und holomorph auf eine Umgebung V des Nullpunktes im C^n abgebildet. U sei ein in V gelegener Hyperzylinder $\{x = (x_1, \ldots, x_n) : |x_\nu| < \varepsilon', \nu = 1, \ldots, n\}$. Für $x \in U$ geht die Abbildung (18) über in:

$${}^{\varrho}\mathfrak{R} : x \to {}^{\varrho}x := \psi \circ {}^{\varrho}\mathfrak{H} \circ \psi^{-1}(x) = {}^{\varrho}D \circ x . \tag{20}$$

Setzen wir $X := \varphi^{-1}(U)$, wobei $\varphi := \psi \circ R \circ \hat{\varphi}$, so stellt (X, φ) eine Karte von X dar, die wieder die Bedingung 1 von Lemma 2 erfüllt. Außerdem haben die Automorphismen $\varphi \circ {}^{\varrho}H \circ \varphi^{-1}$ von $\varphi(X)$ die Form:

(21) $$ {}^{\varrho}\mathfrak{R} : x \to {}^{\varrho}x = {}^{\varrho}D \circ x , $$

was in Komponentenschreibweise folgendermaßen aussieht:

(22) $$ x_\nu \to {}^{\varrho}x_\nu = x_\nu \alpha_r^{p_\nu \varrho} , \quad \nu = 1, \ldots, n . $$

Um die übrigen Bedingungen von Lemma 2 nachzuweisen, erinnern wir uns, daß die Kurve $C_Q : y \to Q^* = H_y(Q)$, $|y| = 1$, (Q ist ein fester Punkt der Klasse K_r) nur Punkte der Klasse K_r enthält, die somit alle von den Transformationen der Gruppe G_r festgelassen werden. Es gilt also:

(23) $$ {}^{\varrho}H \circ H_y(Q) = H_y(Q) . $$

In den Koordinaten x von $U = \varphi(X)$ läßt sich die Kurve folgendermaßen darstellen: $\mathfrak{C}_Q : y \to x' = \mathfrak{R}_y(0)$, wobei $\mathfrak{R}_y(x) := \varphi \circ H_y \circ \varphi^{-1}(x)$ ist. Wir behaupten nun zunächst, daß der Vektor $\left.\frac{d\mathfrak{R}_y(0)}{dy}\right|_{y=1} \neq 0$ ist. Wählt man y^* und y aus einer genügend kleinen Umgebung von 1, so gilt: $\mathfrak{R}_{y^*} \circ \mathfrak{R}_y(x) = \mathfrak{R}_{y^* y}(x)$. Für $x = 0$ erhalten wir hieraus:

(24) $$ \mathfrak{R}_{y^*} \circ \mathfrak{R}_y(0) = \mathfrak{R}_{y^* y}(0) . $$

Bezeichnen wir die Komponenten von $\mathfrak{R}_y(x)$ mit $k_\nu(x, y)$, $\nu = 1, \ldots, n$, und differenzieren wir Gleichung (24) nach y an der Stelle $y = 1$, so ergibt sich:

$$ \sum_{\mu=1}^{n} \left.\frac{\partial \mathfrak{R}_{y^*}(x)}{\partial x_\mu}\right|_{x=0} \left.\frac{d k_\mu(0, y)}{dy}\right|_{y=1} = y^* \left.\frac{d\mathfrak{R}_{y'}(0)}{dy'}\right|_{y'=y^*} $$

oder

(25) $$ \frac{d\mathfrak{R}_{y^*}(0)}{dy^*} = \frac{1}{y^*} M(y^*) \left.\frac{d\mathfrak{R}_y(0)}{dy}\right|_{y=1} , $$

wobei $M(y^*)$ eine Matrix mit den Spaltenvektoren $\left.\frac{\partial \mathfrak{R}_{y^*}(x)}{\partial x_\mu}\right|_{x=0}$, $\mu = 1, \ldots, n$, ist. Nehmen wir nun an, daß $\left.\frac{\partial \mathfrak{R}_y(0)}{\partial y}\right|_{y=1} = 0$ ist, so gilt auch $\frac{d\mathfrak{R}_{y^*}(0)}{dy^*} = 0$ für alle y^* aus einer hinreichend kleinen Umgebung von 1. Das bedeutet aber, daß $\mathfrak{R}_{y^*}(0)$ von y^* unabhängig ist und Q somit ein Fixpunkt der Gruppe L^1 ist, während wir im Widerspruch hierzu angenommen hatten, daß Q zur Klasse K_r, $r > 0$, gehört. Es gilt demnach:

(26) $$ \left.\frac{d\mathfrak{R}_y(0)}{dy}\right|_{y=1} \neq 0 . $$

Ohne Beschränkung der Allgemeinheit können wir annehmen, daß $\left.\frac{d k_\nu(0, y)}{dy}\right|_{y=1} \neq 0$ für $\nu \leqq m$ und $\left.\frac{d k_\nu(0, y)}{dy}\right|_{y=1} = 0$ für $\nu > m$, wobei $1 \leqq m \leqq n$ ist. Damit ist Bedingung 3 von Lemma 2 erfüllt. Es bleibt noch zu zeigen, daß der g. g. T. von $r, p_2, \ldots, p_n$ gleich eins ist. Zu diesem Zweck überlegen wir uns, daß Gleichung (23) in den Koordinaten x von $\varphi(X)$ sich folgendermaßen ausdrücken läßt:

(27) $$ {}^{\varrho}\mathfrak{R} \circ \mathfrak{R}_y(0) = \mathfrak{R}_y(0) . $$

Differentiieren wir Gleichung (27) nach y an der Stelle $y = 1$, so ergibt sich:
$$^{\varrho}D \circ \frac{d\mathfrak{R}_{\nu}(0)}{dy}\bigg|_{y=1} = \frac{d\mathfrak{R}_{\nu}(0)}{dy}\bigg|_{y=1} \text{ oder}$$

$$\alpha_{\nu}^{p_{\nu}\varrho} \frac{dk_{\nu}(0,y)}{dy}\bigg|_{y=1} = \frac{dk_{\nu}(0,y)}{dy}\bigg|_{y=1}. \tag{28}$$

Da $\dfrac{dk_{\nu}(0,y)}{dy}\bigg|_{y=1} \neq 0$ ist für $\nu \leq m$, so kann Gleichung (28) nur erfüllt sein, wenn gilt: $p_1 = \cdots = p_m = 0$. Wie wir sahen, ist der g. g. T. von $r, p_1, \ldots, p_n$ gleich eins. Da aber $p_1 = 0$ ist, so folgt hieraus, daß auch der g. g. T. von $r, p_2, \ldots, p_n$ gleich eins ist. Damit ist Lemma 2 bewiesen.

§ 3. Verallgemeinerung des Cartanschen Abbildungssatzes

Wir wollen uns nun dem Beweis von Satz 1 zuwenden. Wie wir sehen werden, kann der Beweis sehr rasch geführt werden, wenn man den folgenden lokalen Abbildungssatz zur Verfügung hat.

Satz 3: *X^n sei eine k_1-Mannigfaltigkeit. Dann läßt sich jedem Punkt $Q \in X^n$ eine offene Umgebung X_Q und eine komplexe Karte (X_Q, φ_Q) so zuordnen, daß gilt:*

1) *X_Q ist invariant gegenüber den Transformationen H_y, $|y| = 1$, von L^1.*

2) *$\varphi_Q(X_Q)$ ist ein $(p_1^{(Q)}, \ldots, p_n^{(Q)})$-Gebiet im C^n.*

3) *φ_Q ist verträglich mit den Automorphismen H_y von X_Q und den Automorphismen $^{Q}A_y := A_y(p_1^{(Q)}, \ldots, p_n^{(Q)})$ von $\varphi_Q(X_Q)$, d. h. es gilt: $^{Q}A_y \circ \varphi_Q = \varphi_Q \circ H_y$.*

Zum *Beweis von Satz 1* haben wir jetzt nur nachzuweisen, daß $\{(X_Q, \varphi_Q), Q \in X^n\}$ ein komplexer Atlas von X^n ist, der den Bedingungen 1) und 2) von Definition 3 genügt.

Ad 1) ist zu zeigen, daß $\varphi_Q(X_Q \cap X_S)$ wieder ein $(p_1^{(Q)}, \ldots, p_n^{(Q)})$-Gebiet ist. Es gilt: $^{Q}A_y \circ \varphi_Q(X_Q \cap X_S) = \varphi_Q \circ H_y(X_Q \cap X_S) = \varphi_Q(X_Q \cap X_S)$.

Ad 2) haben wir die Gültigkeit der Gleichung $^{Q}A_y \circ \varphi_{QS} = \varphi_{QS} \circ {}^{S}A_y$ nachzuweisen, wobei $\varphi_{QS} := \varphi_Q \circ \varphi_S^{-1}$. Es gilt:

$$\begin{aligned} ^{Q}A_y \circ \varphi_{QS} &= {}^{Q}A_y \circ \varphi_Q \circ \varphi_S^{-1} = \varphi_Q \circ H_y \circ \varphi_S^{-1} = \varphi_Q \circ (\varphi_S \circ H_{y^{-1}})^{-1} \\ &= \varphi_Q \circ ({}^{S}A_{y^{-1}} \circ \varphi_S)^{-1} = \varphi_Q \circ \varphi_S^{-1} \circ {}^{S}A_y = \varphi_{QS} \circ {}^{S}A_y, \text{ q. e. d.} \end{aligned}$$

Beweis von Satz 3: Wir wollen den Beweis zuerst für einen Punkt Q der Klasse K_r, $r > 0$, durchführen. (X, φ) sei eine Karte einer Umgebung $X \subset X^n$ von Q mit den Eigenschaften von Lemma 2. Dabei kann angenommen werden (siehe Beweis von Lemma 2), daß $\varphi(X)$ ein Hyperzylinder $\{x : |x_{\nu}| < \varepsilon', \nu = 1, \ldots, n\}$ ist. Auf Grund von Lemma 1 läßt sich die Abbildung H: $(P, y) \to P^* = H_y(P)$ von $X \times \{y : |y| = 1\}$ in X^n zu einer holomorphen Abbildung $H^* | X \times \{y : 1 - \varepsilon < |y| < 1 + \varepsilon\} \to X^n$ fortsetzen, wobei ε eine passend gewählte positive Zahl ist. Die Abbildung H^* sei wieder in der Form H^*: $(P, y) \to P^* = H_y(P)$ geschrieben.

$$K : (x, y) \to P^* := H_y \circ \varphi^{-1}(x) \tag{29}$$

stellt dann eine holomorphe Abbildung von $\varphi(X)\times\{y: 1-\varepsilon<|y|<1+\varepsilon\}$ in X^n dar. Die Beschränkung von K auf $R(\varepsilon',\varepsilon):=\{(x,y):|x_\nu|<\varepsilon',\nu=2,\ldots,n,$ $x_1=0,\ 1-\varepsilon<|y|<1+\varepsilon\}$ sei mit

$$\hat{K}:(x,y)\to P^*=\hat{K}(x,y) \tag{30}$$

bezeichnet. Wir behaupten, daß die Abbildung (30) in einer hinreichend kleinen Umgebung von $x=0$, $y=1$ eineindeutig ist. Zu diesem Zweck haben wir nur nachzuweisen, daß die Determinante

$$D:=\operatorname{Det}\left(\frac{\partial\mathfrak{K}_y(x)}{\partial y},\frac{\partial\mathfrak{K}_y(x)}{\partial x_2},\ldots,\frac{\partial\mathfrak{K}_y(x)}{\partial x_n}\right)\Bigg|_{x=0,\,y=1}$$

ungleich Null ist. Dabei ist $\mathfrak{K}_y(x)=\varphi\circ K(x,y)$ wie in Lemma 2 definiert. Da $x\to x'=\mathfrak{K}_y(x)$ für $y=1$ die identische Abbildung darstellt, so gilt: $\left.\frac{\partial\mathfrak{K}_y(x)}{\partial x_\nu}\right|_{x=0,\,y=1}=(\delta_{\nu\mu}),\ \nu=1,\ldots,n$. Somit ist $D=\left.\frac{\partial k_1(x,y)}{\partial y}\right|_{x=0,\,y=1}$. Auf Grund von Lemma 2 ist dann $D\neq 0$. Für genügend kleine positive Werte von ε', ε und δ stellt (30) somit eine eineindeutige holomorphe Abbildung von $R(\varepsilon',\varepsilon,\delta):=\{(x,y):(x,y)\in R(\varepsilon',\varepsilon),|\arg(y)|<\delta\}$ in X^n dar. Für $r>1$ stellt (30) sicher keine eineindeutige Abbildung von $R(\varepsilon',\varepsilon)$ in X^n dar, denn es gilt: $K({}^\varrho D\circ x, y\alpha_r^{-\varrho})=H_{y\alpha_r^{-\varrho}}\circ\varphi^{-1}({}^\varrho D\circ x)=H_y\circ{}^{-\varrho}H\circ\varphi^{-1}({}^\varrho D\circ x)$ $=(H_y\circ\varphi^{-1})\circ(\varphi\circ{}^{-\varrho}H\circ\varphi^{-1})({}^\varrho D\circ x)=H_y\circ\varphi^{-1}\circ{}^{-\varrho}\mathfrak{R}({}^\varrho D\circ x)$ $=H_y\circ\varphi^{-1}(x)=K(x,y)$. Da mit $(x,y)\in R(\varepsilon',\varepsilon)$ auch $({}^\varrho D\circ x,y\alpha_r^{-\varrho})\in R(\varepsilon',\varepsilon)$, so folgt:

$$\hat{K}({}^\varrho D\circ x,y\alpha_r^{-\varrho})=\hat{K}(x,y)\,. \tag{31}$$

Die Punkte $({}^\varrho D\circ x,y\alpha_r^{-\varrho})\in R(\varepsilon',\varepsilon)$, $\varrho=0,1,\ldots,r-1$, werden also durch (30) auf denselben Punkt in X^n abgebildet. Wir behaupten nun, daß bei der Abbildung (30) nur die Punkte $({}^\varrho D\circ x,y\alpha_r^{-\varrho})$, $\varrho=1,\ldots,r-1$, denselben Bildpunkt wie (x,y) besitzen. Wir zeigen zunächst, daß die Behauptung für $\hat{R}(\varepsilon',\varepsilon,\delta):=\bigcup_{\varrho=0}^{r-1}\{(x,y):(x,y)\in R(\varepsilon',\varepsilon),|\arg(y)-\varrho\,2\pi/r|<\delta\}$ richtig ist. Wir wollen annehmen, daß $\hat{K}(x,y)=\hat{K}(x',y')$ ist, wobei (x,y) und $(x',y')\in$ $\in\hat{R}(\varepsilon',\varepsilon,\delta)$. Dann gibt es ganze Zahlen ϱ und ϱ', so daß $({}^\varrho D\circ x,y\alpha_r^{-\varrho})$ und $({}^{\varrho'}D\circ x',y'\alpha_r^{-\varrho'})$ in $R(\varepsilon',\varepsilon,\delta)$ liegen. Auf Grund von Relation (31) gilt: $\hat{K}({}^\varrho D\circ x,y\alpha_r^{-\varrho})=\hat{K}({}^{\varrho'}D\circ x',y'\alpha_r^{-\varrho'})$. Da durch (30) das Gebiet $R(\varepsilon',\varepsilon,\delta)$ eineindeutig in X^n abgebildet wird, so ist $({}^\varrho D\circ x,y\alpha_r^{-\varrho})=({}^{\varrho'}D\circ x',y'\alpha_r^{-\varrho'})$, oder anders ausgedrückt: $(x',y')=({}^{\varrho^*}D\circ x,y\alpha_r^{-\varrho^*})$, wobei $\varrho^*=\varrho-\varrho'$ ist, q. e. d. Wir wollen jetzt einmal annehmen, daß die obige Behauptung für $(x,y)\in R(\varepsilon',\varepsilon)$, wie klein ε' und ε auch gewählt sein mögen, falsch ist. Dann gibt es Folgen $\{\varepsilon_\varkappa\}$ und $\{\varepsilon'_\varkappa\}$, $\varkappa=1,2,\ldots$, wobei $\varepsilon\geqq\varepsilon_\varkappa>0$, $\varepsilon'\geqq\varepsilon'_\varkappa>0$, $\lim_{\varkappa\to\infty}\varepsilon_\varkappa=\lim_{\varkappa\to\infty}\varepsilon'_\varkappa=0$, und es gibt Punkte $({}^\varkappa x,{}^\varkappa y)$, $({}^\varkappa x^*,{}^\varkappa y^*)\in R(\varepsilon'_\varkappa,\varepsilon_\varkappa)$, so daß gilt: $\hat{K}({}^\varkappa x,{}^\varkappa y)=\hat{K}({}^\varkappa x^*,{}^\varkappa y^*)$, $({}^\varkappa x^*,{}^\varkappa y^*)\neq({}^\varrho D\circ{}^\varkappa x,{}^\varkappa y\alpha_r^{-\varrho})$ für $\varrho=0,1,\ldots,r-1$. Zunächst einmal macht man sich klar, daß $({}^\varkappa x,{}^\varkappa y)$ stets so gewählt werden kann, daß $\arg({}^\varkappa y)=0$ ist. Denn ist $\arg({}^\varkappa y)=\alpha$, so setzen wir ${}^\varkappa\hat{y}:={}^\varkappa y e^{-i\alpha}$ und ${}^\varkappa\hat{y}^*:={}^\varkappa y^* e^{-i\alpha}$. $({}^\varkappa x,{}^\varkappa\hat{y})$ und $({}^\varkappa x^*,{}^\varkappa\hat{y}^*)$ liegen wieder in $R(\varepsilon'_\varkappa,\varepsilon_\varkappa)$, und es gilt:

$$\hat{K}({}^\varkappa x,{}^\varkappa\hat{y})=\hat{K}({}^\varkappa x^*,{}^\varkappa\hat{y}^*)\,, \tag{32}$$

während $({}^{\varkappa}x^*, {}^{\varkappa}\hat{y}^*) \neq ({}^{\varrho}D \circ {}^{\varkappa}x, {}^{\varkappa}\hat{y}\,\alpha_r^{-\varrho})$ für $\varrho = 0, 1, \ldots, r-1$. Da $({}^{\varkappa}x, {}^{\varkappa}\hat{y}) \in$ $\in R(\varepsilon', \varepsilon, \delta)$, so müssen wir annehmen, daß $({}^{\varkappa}x^*, {}^{\varkappa}\hat{y}^*)$ nicht zu $\hat{R}(\varepsilon', \varepsilon, \delta)$ gehört, d. h. es gilt: $|\arg({}^{\varkappa}\hat{y}^*) - \varrho\, 2\pi/r| \geq \delta > 0$ für $\varrho = 0, 1, \ldots, r-1$. δ ist dabei eine von $\varkappa$ unabhängige positive reelle Zahl. Es gilt: $\lim\limits_{\varkappa \to \infty} {}^{\varkappa}x = \lim\limits_{\varkappa \to \infty} {}^{\varkappa}x^* = 0$, $\lim {}^{\varkappa}\hat{y} = 1$ und (evtl. nach Auswahl einer Teilfolge erst) $\lim\limits_{\varkappa \to \infty} {}^{\varkappa}\hat{y}^* = e^{i\beta}$, wo β eine reelle Zahl ist, die den Bedingungen $0 \leq \beta \leq 2\pi$ und $|\beta - \varrho\, 2\pi/r| \geq \delta > 0$, $\varrho = 0, \ldots, r-1$, genügt. Aus (32) folgt nun durch Limesbildung: $\hat{K}(0, 1)$ $= \hat{K}(0, e^{i\beta})$, oder anders ausgedrückt: $Q = H_{e^{i\beta}}(Q)$. Da Q aus der Klasse K_r, $r > 0$, ist, so kann $e^{i\beta}$ nur eine der Zahlen α_r^{ϱ}, $\varrho = 0, \ldots, r-1$, sein, was im Widerspruch zu $|\beta - \varrho\, 2\pi/r| \geq \delta > 0$, $\varrho = 0, \ldots, r-1$, steht. Somit gibt es also positive reelle Zahlen ε' und ε, so daß Abbildung (30) die Eigenschaft besitzt, daß nur die Punkte $({}^{\varrho}D \circ x, y\alpha_r^{-\varrho})$, $\varrho = 1, \ldots, r-1$, denselben Bildpunkt wie (x, y) haben. Im Falle $r = 1$ ist die Abbildung (30) von $R(\varepsilon', \varepsilon)$ in X^n somit schon eineindeutig.

Durch die Zuordnung:

$$(33) \qquad \chi : (x_1, \ldots, x_n, y) \to (z_1, \ldots, z_n) = (y, x_2, \ldots, x_n)$$

wird $R(\varepsilon', \varepsilon)$ eineindeutig und holomorph auf ein Hartogssches Gebiet $\hat{U} := \{z : 1 - \varepsilon < |z_1| < 1 + \varepsilon, |z_\nu| < \varepsilon', \nu = 2, \ldots, n\}$ abgebildet.

$$(34) \qquad F : z \to P^* := \hat{K} \circ \chi^{-1}(z)$$

stellt somit eine holomorphe Abbildung von $\hat{U}$ in X^n dar, die der Bedingung genügt, daß genau die Punkte ${}^{\varrho}z := (z_1\alpha_r^{-\varrho}, z_2\alpha_r^{p_2\varrho}, \ldots, z_n\alpha_r^{p_n\varrho})$, $\varrho = 1, \ldots, r-1$, denselben Bildpunkt wie $z = (z_1, \ldots, z_n)$ haben. Wir wollen annehmen, daß $\varepsilon < 1$ ist, dann liegen die Punkte ${}^{\varrho}z$, $\varrho = 0, \ldots, r-1$, alle isoliert, und die Abbildung (34) ist damit lokal eineindeutig.

Eine zweite lokal eineindeutige holomorphe Abbildung von $\hat{U}$ auf ein $(p_1, \ldots, p_n)$-Gebiet $U_Q \subset C^n$ ist durch die folgende Zuordnung gegeben:

$$(35) \qquad F^* : (z_1, \ldots, z_n) \to (w_1, \ldots, w_n) = (z_1^{p_1}, z_2 z_1^{p_2}, \ldots, z_n z_1^{p_n}),$$

wobei $p_1 = r$ ist. Wie bei Abbildung (34) gehen auch hier genau die Punkte ${}^{\varrho}z$, $\varrho = 1, \ldots, r-1$, in denselben Bildpunkt wie z über.

Damit liefert dann:

$$(36) \qquad \varphi_Q : P^* \to w := F^* \circ F^{-1}(P)$$

eine eineindeutige holomorphe Abbildung von $X_Q := F(\hat{U}) \subset X^n$ auf das $(p_1, \ldots, p_n)$-Gebiet U_Q.

Wir haben nun im einzelnen zu prüfen, ob (X_Q, φ_Q) eine Karte der Umgebung X_Q von Q mit den Eigenschaften 1)—3) von Satz 2 ist.

Ad 1): Ein Punkt $P \in X_Q$ kann geschrieben werden als $P = \hat{K}(x, y)$, $(x, y) \in R(\varepsilon', \varepsilon)$. Um zu zeigen, daß $H_{y^*}(P)$, $|y^*| = 1$, wieder zu X_Q gehört, hat man nur folgendes auszurechnen: $H_{y^*}(P) = H_{y^*} \circ \hat{K}(x, y) = H_{y^*} \circ H_y \circ \varphi^{-1}(x)$ $= H_{y^* y} \circ \varphi^{-1}(x) = \hat{K}(x, y y^*)$, $(x, y y^*) \in R(\varepsilon', \varepsilon)$.

Ad 2): $U_Q := \varphi_Q(X_Q)$ ist per definitionem ein $(p_1, \ldots, p_n)$-Gebiet. Wir haben uns nur zu überlegen, daß der g. g. T. von $p_1, \ldots, p_n$ gleich eins ist. Dies folgt aber sofort aus Lemma 2, da wir $p_1 = r$ gesetzt hatten.

Ad 3): Zu zeigen ist:

$$ {}^{Q}A_{y^*} \circ \varphi_Q(P) = \varphi_Q \circ H_{y^*}(P) . \tag{37}$$

$P \in X_Q$ kann dargestellt werden durch $P = \hat{K}(x, y) = \hat{K} \circ \chi^{-1}(z) = F(z)$. Um die Gültigkeit von Gleichung (37) nachzuweisen, rechnen wir beide Seiten getrennt aus:

$$\begin{aligned} {}^{Q}A_{y^*} \circ F^* \circ F^{-1}(P) &= {}^{Q}A_{y^*} \circ F^*(z) = {}^{Q}A_{y^*} \circ (z_1^{p_1}, z_2 z_1^{p_2}, \ldots, z_n z_1^{p_n}) \\ &= (z_1^{p_1} y^{*p_1}, z_2 z_1^{p_2} y^{*p_2}, \ldots, z_n z_1^{p_n} y^{*p_n}) . \end{aligned}$$

$$\begin{aligned} \varphi_Q \circ H_{y^*}(P) &= \varphi_Q \circ H_{y^*} \circ \hat{K}(x, y) = \varphi_Q \circ \hat{K}(x, y y^*) = F^* \circ F^{-1} \circ \hat{K}(x, y y^*) \\ &= F^* \circ \chi(x, y y^*) = F^*(z_1 y^*, z_2, \ldots, z_n) \\ &= ((z_1 y^*)^{p_1}, z_2 (z_1 y^*)^{p_2}, \ldots, z_n (z_1 y^*)^{p_n}) . \end{aligned}$$

Damit ist Satz 3 für Punkte Q der Klasse K_r, $r \geqq 1$, bewiesen.

Sei nun Q ein Punkt der Klasse K_0 und (X, φ) eine komplexe Karte einer Umgebung X von Q mit der Eigenschaft: $\varphi(Q) = 0$. Wir können stets annehmen, daß $X \subset X^n$ und $\varphi(X)$ ein beschränktes Gebiet im C^n ist. Es gibt dann eine Umgebung $X' \subset X$ von Q, so daß $P^* = H_y(P) \in X$ für alle $(P, y) \in X' \times \{y : |y| = 1\}$. Andernfalls existiert nämlich eine Folge von Punkten $\{(P_\nu, y_\nu) : P_\nu \in X', \; |y_\nu| = 1, \; \nu = 1, 2, \ldots\}$ mit $\lim_{\nu \to \infty} P_\nu = Q$ und $\lim_{\nu \to \infty} y_\nu = y_0$, $|y_0| = 1$, so daß $P_\nu^* = H_{y_\nu}(P_\nu) \notin X$. Das steht aber im Widerspruch zu: $\lim_{\nu \to \infty} P_\nu^* = \lim_{\nu \to \infty} H_{y_\nu}(P_\nu) = H_{y_0}(Q) = Q$. Definieren wir nun $\hat{X} := \{P : P = H_y(P^*), P^* \in X', |y| = 1\}$, $\hat{\varphi} := \varphi | \hat{X}$, so stellt $(\hat{X}, \hat{\varphi})$ eine Karte der Umgebung $\hat{X} \subset X$ von Q dar, wobei gilt: $\hat{\varphi}(Q) = 0$, $\hat{\varphi}(\hat{X})$ ist beschränkt, und $\hat{X}$ ist invariant gegenüber den Transformationen H_y der Gruppe L^1. In den Koordinaten $z = (z_1, \ldots, z_n)$ von $\hat{\varphi}(\hat{X})$ lassen sich diese Transformationen folgendermaßen schreiben:

$$\mathfrak{F}_y : z \to z' := \hat{\varphi} \circ H_y \circ \hat{\varphi}^{-1}(z) , \tag{38}$$

wobei die Komponenten $f_\nu(z, y)$, $\nu = 1, \ldots, n$, von $\mathfrak{f}_y(z)$ die folgende Entwicklung besitzen:

$$f_\nu(z, y) = \sum_{\mu = 1}^{n} a_{\nu\mu}(y) \, z_\mu + \cdots , \tag{39}$$

oder auch:

$$\mathfrak{f}_y(z) = A_y \circ z + \cdots , \tag{40}$$

wobei $A_y := (a_{\nu\mu}(y))$ ist. Da $\hat{\varphi}(\hat{X})$ beschränkt ist, gilt der Cartansche Eindeutigkeitssatz. Die Transformationen (38) sind somit eindeutig durch den Linearteil A_y bestimmt. Es gilt ferner: $A_y \circ A_{y^*} = A_{y y^*}$, d. h. die Matrizen A_y, $|y| = 1$, bilden eine kompakte 1-dimensionale Liesche Gruppe. Es gibt dann eine nicht-singuläre lineare Transformation S, so daß die Matrizen $D_y = S \circ A_y \circ S^{-1}$ Diagonalgestalt besitzen. Die Elemente in der Diagonalen

von D_y seien mit $\delta_\nu(y)$, $\nu = 1, \ldots, n$, bezeichnet. Wegen $D_y \circ D_{y^*} = D_{yy^*}$ gilt auch: $\delta_\nu(y)\, \delta_\nu(y^*) = \delta_\nu(yy^*)$. Daraus folgt: $\delta_\nu(y) \left.\frac{d\,\delta_\nu(y^*)}{d\,y^*}\right|_{y^*=1} = y \frac{d\,\delta_\nu(y)}{d\,y}$ oder

$$(41) \qquad \delta_\nu(y)\, p_\nu = y \frac{d\,\delta_\nu(y)}{d\,y}, \quad \nu = 1, \ldots, n,$$

wobei $p_\nu := \left.\frac{d\,\delta_\nu(y^*)}{d\,y^*}\right|_{y^*=1}$. Als Lösung der Differentialgleichungen (41) mit den Nebenbedingungen $\delta_\nu(1) = 1$ erhalten wir:

$$(42) \qquad \delta_\nu(y) = y^{p_\nu}, \quad \nu = 1, \ldots, n.$$

Dabei sind die p_ν ganze Zahlen. Der g. g. T. von $p_1, \ldots, p_n$ ist gleich eins, da D_y nur für $y = 1$ gleich der Einheitsmatrix ist.

Durch die Koordinatentransformation $z \to w := S \circ z$ erhalten wir eine neue Karte $(\hat{X}, \varphi^*)$ von $\hat{X}$, wobei $\varphi^* := S \circ \hat{\varphi}$ ist. Die Abbildungen (38) gehen dabei über in Automorphismen von $\varphi^*(\hat{X})$:

$$(43) \qquad \mathfrak{H}_y : w \to w' := \varphi^* \circ H_y \circ \varphi^{*-1}(w) = D_y \circ w + \cdots$$

Die holomorphe Abbildung:

$$(44) \qquad \psi : w \to x := \frac{1}{2\pi i} \oint_{|y|=1} D_{y^{-1}}\, \mathfrak{H}_y(w)\, d(\log y) = w + \cdots$$

von $\varphi^*(\hat{X})$ in den C^n ist in einer Umgebung des Nullpunktes eineindeutig. Wir wollen annehmen, daß $\hat{X}$ von vorneherein so klein gewählt war, daß die Abbildung ψ in ganz $\varphi^*(\hat{X})$ eineindeutig ist. Man verifiziert sehr leicht, daß

$$(45) \qquad D_y \circ \psi(w) = \psi \circ \mathfrak{H}_y(w),$$

d. h. $\psi \circ \varphi^*(\hat{X})$ ist ein $(p_1, \ldots, p_n)$-Gebiet.

Setzen wir $\varphi_Q := \psi \circ \varphi^*$, $X_Q := \hat{X}$, so stellt (X_Q, φ_Q) eine Karte der Umgebung X_Q von Q dar, die sämtliche Bedingungen von Satz 3 erfüllt. Es bleibt hierzu nur noch zu zeigen, daß

$$(46) \qquad {}^QA_y \circ \varphi_Q = \varphi_Q \circ H_y.$$

Dabei ist ${}^QA_y$ per definitionem gleich D_y. Setzen wir in Gleichung (45): $w = \varphi^*(P)$, so ergibt sich: $D_y \circ \psi \circ \varphi^*(P) = \psi \circ \mathfrak{H}_y \circ \varphi^*(P) = (\psi \circ \varphi^*) \circ \circ (\varphi^{*-1} \circ \mathfrak{H}_y \circ \varphi^*)(P) = (\psi \circ \varphi^*) \circ H_y(P)$ oder ${}^QA_y \circ \varphi_Q(P) = \varphi_Q \circ H_y(P)$, q. e. d. Damit ist Satz 3 auch für Punkte der Klasse K_0 bewiesen.

Literatur

[1] Behnke, H., u. E. Peschl: Zur Theorie der Funktionen mehrerer komplexer Veränderlicher. Der Cartansche Eindeutigkeitssatz in unbeschränkten Körpern. Math. Ann. **114**, 69—73 (1937). — [2] Behnke, H., u. P. Thullen: Theorie der Funktionen mehrerer komplexer Veränderlicher. Ergebn. Math. **3** (1934). — [3] Cartan, H.: Les fonctions de deux variables complexes et le problème de la représentation analytique. J. Math. pures et appl. **9**, 10, 1—114 (1931). — [4] Chevalley, C.: Theory of Lie groups I. Princeton Univ. Press 1946. — [5] Remmert, R.: Holomorphe und meromorphe Abbildungen komplexer Räume. Math. Ann. **133**, 328—370 (1957).

(Eingegangen am 18. April 1959)

HORNFECK, B.
Math. Annalen 138, 442—449 (1959)

Über natürliche Zahlen, deren Primteiler in mindestens k-ter Potenz auftreten

Von

BERNHARD HORNFECK in Braunschweig

Im Verlauf zahlentheoretischer Untersuchungen ergibt sich bisweilen die Frage nach der Anzahlfunktion

$$A_k(x) = \sum_{\substack{a_i \leq x \\ a_i \in \mathfrak{A}_k}} 1 \qquad (k \geq 1, \text{ganz})$$

der Menge $\mathfrak{A}_k$ aller natürlichen Zahlen, deren Primteiler alle in mindestens k-ter Potenz auftreten. Bekannt ist (vgl. § 1, Hilfssatz 1)

$$A_k(x) = O\left(x^{\frac{1}{k}}\right) \qquad (x \to \infty);$$

eine *asymptotische* Abschätzung von $A_k(x)$ scheint jedoch noch nicht vorzuliegen. Sie wird in § 2, Satz 3 angegeben.

Es ist das Ziel dieser Untersuchung, die angeschnittene Frage in einem allgemeineren Rahmen zu beantworten. Wir betrachten die Menge $\mathfrak{M}_k$ aller derjenigen natürlichen Zahlen aus $\mathfrak{A}_k$, deren sämtliche Primteiler aus einer fest gewählten Teilmenge $\mathfrak{T}$ der Menge aller Primzahlen stammen; über die Anzahlfunktion $T(x)$ von $\mathfrak{T}$ setzen wir die asymptotische Beziehung

$$T(x) \sim \tau \frac{x}{\log x}, \quad \tau > 0, \tag{1}$$

voraus.

In § 2, Satz 1 wird dann die Anzahlfunktion $M_k(x)$ von $\mathfrak{M}_k$ asymptotisch abgeschätzt; die Sätze 2 und 3 ergeben sich als Spezialisierungen dieses allgemeinen Ergebnisses. Es stellt sich heraus, daß $M_k(x)$ nur unter der notwendigen aber nicht zugleich auch hinreichenden Bedingung $\tau = 1$ von der Größenordnung $x^{\frac{1}{k}}$ sein kann. Satz 2 gibt die asymptotische Abschätzung von $M_1(x)$; er ist erstmals von WIRSING [6] bewiesen worden. Die vorliegende Untersuchung benutzt eine Erweiterung der gleichen Beweisidee, die auf der Anwendung eines reellen Taubersatzes von HARDY und LITTLEWOOD [2] (§ 1, Hilfssatz 7) beruht.

§ 1. (Hilfssätze, Definitionen)

Aus beweistechnischen Gründen soll für alles Folgende $1 \in \mathfrak{T}$, also auch $1 \in \mathfrak{M}_k$, angenommen werden.

Hilfssatz 1: *Es gilt*

$$M_k(x) = O\left(x^{\frac{1}{k}}\right) \qquad (x \to \infty). \tag{2}$$

Beweis: Jedes $m \in \mathfrak{M}_k$ läßt sich in der Gestalt

$$m = n_0^k n_1^{k+1} \ldots n_{k-1}^{2k-1}$$

mit natürlichen Zahlen n_i $(i = 0, 1, \ldots, k-1)$ schreiben. Die Anzahl aller $n_0^k n_1^{k+1} \leq x$ ist aber höchstens

$$\sum_{n_1^{k+1} \leq x} \left(\frac{x}{n_1^{k+1}}\right)^{\frac{1}{k}} < x^{\frac{1}{k}} \sum_{n_1=1}^{\infty} \frac{1}{n_1^{1+\frac{1}{k}}} = O\left(x^{\frac{1}{k}}\right),$$

und $(k-2)$-malige Wiederholung dieser Rechnung liefert (2).

Hilfssatz 2: *Es ist*

$$\sum_{\substack{p \leq x \\ p \in \mathfrak{T}}} \frac{1}{p} \sim \tau \log\log x \qquad (x \to \infty).$$

Außerdem gilt für jedes $u > 1$

$$\sum_{\substack{e^x < p \leq e^{ux} \\ p \in \mathfrak{T}}} \frac{1}{p} = \tau \log u + o(1) \qquad (x \to \infty). \tag{3}$$

Beweis: Man hat mit (1)

$$\sum_{\substack{p \leq x \\ p \in \mathfrak{T}}} \frac{1}{p} = \int_{1-}^{x} \frac{d\,T(t)}{t} = \frac{T(x)}{x} + \int_{1}^{x} \frac{T(t)}{t^2}\,dt$$

$$= \tau \log\log x + o(\log\log x) \qquad (x \to \infty).$$

Genauso wird

$$\sum_{\substack{e^x < p \leq e^{ux} \\ p \in \mathfrak{T}}} \frac{1}{p} = \int_{e^x}^{e^{ux}} \frac{d\,T(t)}{t} = \frac{T(e^{ux})}{e^{ux}} - \frac{T(e^x)}{e^x} + \int_{e^x}^{e^{ux}} \frac{T(t)}{t^2}\,dt$$

$$= o(1) + \int_{e^x}^{e^{ux}} \frac{\tau + o(1)}{t \log t}\,dt = \tau \log u + o(1) \qquad (x \to \infty).$$

Definitionen: *Wir setzen*

$$l_k(x) =_{Df} \sum_{\substack{m \leq x \\ m \in \mathfrak{M}_k}} \log m,$$

$$\vartheta_k(x) =_{Df} \sum_{\substack{p \leq x^{\frac{1}{k}} \\ p \in \mathfrak{T}}} \log p.$$

Hilfssatz 3: *Es ist* $(x \to \infty)$

$$l_k(x) = k \sum_{\substack{m \leq x \\ m \in \mathfrak{M}_k}} \vartheta_k\left(\frac{x}{m}\right) + O\left(x^{\frac{1}{k}}\right). \tag{4}$$

Beweis: Wir schreiben

$$l_k(x) = \sum_{\substack{m \leq x \\ m \in \mathfrak{M}_k}} \log m = \sum_{\substack{m \leq x \\ m \in \mathfrak{M}_k}} \sum_{\substack{p^\alpha | m \\ p^{\alpha+1} \nmid m \\ p \in \mathfrak{T}}} \log p^\alpha$$

$$= \sum_{\substack{m' \in \mathfrak{M}_k, p \in \mathfrak{T}, \alpha \geq k \\ m' p^\alpha \leq x \\ (m', p) = 1}} \log p^\alpha = \sum_{\substack{m' \in \mathfrak{M}_k, p \in \mathfrak{T}, \alpha \geq k \\ m' p^\alpha \leq x}} \log p^\alpha - \sum_{\substack{m' \in \mathfrak{M}_k, p \in \mathfrak{T}, \alpha \geq k \\ p | m' \\ m' p^\alpha \leq x}} \log p^\alpha$$

mit

$$\sum_{\substack{m'\in\mathfrak{M}_k,\, p\in\mathfrak{T},\, \alpha\ge k\\ p\mid m'\\ m'p^\alpha\le x}} \log p^\alpha = \sum_{\substack{m''\in\mathfrak{M}_k,\, p\in\mathfrak{T},\, \gamma\ge 2k\\ (p,m'')=1\\ m''p^\gamma\le x}} \log p \sum_{\substack{\alpha\ge k,\, \beta\ge k\\ \alpha+\beta=\gamma}} \alpha < \sum_{\substack{m''\in\mathfrak{M}_k,\, p\in\mathfrak{T},\, \gamma\ge 2k\\ (p,m'')=1\\ m''p^\gamma\le x}} \gamma^2 \log p$$

$$\le \sum_{p\in\mathfrak{T},\, \gamma\ge 2k} \gamma^2 M_k\left(\frac{x}{p^\gamma}\right)\log p\,;$$

Anwendung von Hilfssatz 1 liefert weiter

$$\sum_{\substack{m'\in\mathfrak{M}_k,\, p\in\mathfrak{T},\, \alpha\ge k\\ p\mid m'\\ m'p^\alpha\le x}} \log p^\alpha \le c_1 x^{\frac{1}{k}} \sum_{p\in\mathfrak{T},\, \gamma\ge 2k} \frac{\gamma^2}{p^{\frac{\gamma}{k}}}\log p = c_1 x^{\frac{1}{k}} \sum_{p\in\mathfrak{T}} \log p \sum_{\gamma\ge 2k} \frac{\gamma^2}{p^{\frac{\gamma}{k}}}$$

$$\le c_1 x^{\frac{1}{k}} \sum_{p\in\mathfrak{T}} \log p\cdot\frac{c_2}{p^2} = O\left(x^{\frac{1}{k}}\right) \qquad (x\to\infty)\,.$$

Also ist

$$l_k(x) = \sum_{\substack{m'\in\mathfrak{M}_k,\, p\in\mathfrak{T},\, \alpha\ge k\\ m'p^\alpha\le x}} \log p^\alpha + O\left(x^{\frac{1}{k}}\right)$$

$$= \sum_{\substack{m'\in\mathfrak{M}_k,\, p\in\mathfrak{T}\\ m'p^k\le x}} k\log p + \sum_{\substack{m'\in\mathfrak{M}_k,\, p\in\mathfrak{T},\, \alpha\ge k+1\\ m'p^\alpha\le x}} \alpha\log p + O\left(x^{\frac{1}{k}}\right)$$

$$= k\sum_{\substack{m'\le x\\ m'\in\mathfrak{M}_k}} \sum_{\substack{p\le\left(\frac{x}{m'}\right)^{\frac{1}{k}}\\ p\in\mathfrak{T}}} \log p + R(x) + O\left(x^{\frac{1}{k}}\right)$$

$$= k\sum_{\substack{m'\le x\\ m'\in\mathfrak{M}_k}} \vartheta_k\left(\frac{x}{m'}\right) + R(x) + O\left(x^{\frac{1}{k}}\right) \qquad (x\to\infty)\,,$$

und wir haben noch $R(x) = O\left(x^{\frac{1}{k}}\right)$ zu zeigen:

$$R(x) = \sum_{p\in\mathfrak{T}} \log p \sum_{\alpha\ge k+1} \alpha M_k\left(\frac{x}{p^\alpha}\right) \le c_1 x^{\frac{1}{k}} \sum_{p\in\mathfrak{T}} \log p \sum_{\alpha\ge k+1} \frac{\alpha}{p^{\frac{\alpha}{k}}}$$

$$\le c_1 x^{\frac{1}{k}} \sum_{p\in\mathfrak{T}} \log p\cdot\frac{c_3}{p^{1+\frac{1}{k}}} = O\left(x^{\frac{1}{k}}\right) \qquad (x\to\infty)\,.$$

Hilfssatz 4: *Es ist*

(5) $$\vartheta_k(x) \sim \tau x^{\frac{1}{k}} \qquad (x\to\infty)\,.$$

Beweis: Nach (1) ist

$$\vartheta_k(x) = \int_{1-}^{x^{\frac{1}{k}}} \log t\; dT(t) = \log x^{\frac{1}{k}}\; T\left(x^{\frac{1}{k}}\right) - \int_{1}^{x^{\frac{1}{k}}} \frac{T(t)}{t}\,dt = \tau x^{\frac{1}{k}} + o\left(x^{\frac{1}{k}}\right) \qquad (x\to\infty)\,.$$

Definition: *Wir setzen*

$$m_k(x) = \sum_{\substack{m\le x\\ m\in\mathfrak{M}_k}} \frac{1}{m^{\frac{1}{k}}}\,.$$

Hilfssatz 5: *Es ist*

(6) $$M_k(x) \sim k\tau\, m_k(x)\cdot\frac{x^{\frac{1}{k}}}{\log x} \qquad (x\to\infty)\,.$$

Beweis: Nach Hilfssatz 3 ist $(x \to \infty)$

(7) $$l_k(x) - k\tau m_k(x)\, x^{\frac{1}{k}} = k \sum_{\substack{m \leq x \\ m \in \mathfrak{M}_k}} \vartheta_k\left(\frac{x}{m}\right) - k\tau m_k(x)\, x^{\frac{1}{k}} + O\left(x^{\frac{1}{k}}\right);$$

hierin wird nach Eintragen von (5)

$$k \sum_{\substack{m \leq x \\ m \in \mathfrak{M}_k}} \vartheta_k\left(\frac{x}{m}\right) - k\tau m_k(x) \cdot x^{\frac{1}{k}}$$

$$= k \sum_{\substack{m \leq x \\ m \in \mathfrak{M}_k}} \left\{\tau \frac{x^{\frac{1}{k}}}{m^{\frac{1}{k}}} + o\left(\frac{x^{\frac{1}{k}}}{m^{\frac{1}{k}}}\right)\right\} - k\tau m_k(x)\, x^{\frac{1}{k}} = \sum_{\substack{m \leq x \\ m \in \mathfrak{M}_k}} o\left(\frac{x^{\frac{1}{k}}}{m^{\frac{1}{k}}}\right) \qquad \left(\frac{x}{m} \to \infty\right),$$

und wegen $m_k(x) \geqq \sum_{\substack{p \leq x \\ p \in \mathfrak{T}}} \frac{1}{p} \to \infty$ für $x \to \infty$ (Hilfssatz 2) bekommen wir weiter

$$k \sum_{\substack{m \leq x \\ m \in \mathfrak{M}_k}} \vartheta_k\left(\frac{x}{m}\right) - k\tau m_k(x)\, x^{\frac{1}{k}} = \sum_{\substack{\frac{x}{m_k(x)} \leq m \leq x \\ m \in \mathfrak{M}_k}} o\left(\frac{x^{\frac{1}{k}}}{m^{\frac{1}{k}}}\right) + \sum_{\substack{\frac{x}{m} > m_k(x) \\ m \in \mathfrak{M}_k}} o\left(\frac{x^{\frac{1}{k}}}{m^{\frac{1}{k}}}\right)$$

$$= O\left(x^{\frac{1}{k}} \sum_{\frac{x}{m_k(x)} \leq n \leq x} \frac{1}{n^{\frac{1}{k}}}\right) + o\left(\sum_{\substack{m < \frac{x}{m_k(x)} \\ m \in \mathfrak{M}_k}} \frac{x^{\frac{1}{k}}}{m^{\frac{1}{k}}}\right)$$

$$= O\left(x^{\frac{1}{k}} \log m_k(x)\right) + o\left(x^{\frac{1}{k}}\, m_k(x)\right) = o\left(x^{\frac{1}{k}}\, m_k(x)\right) \qquad (x \to \infty).$$

In (7) eingesetzt, kommt

(8) $$l_k(x) \sim k\tau m_k(x)\, x^{\frac{1}{k}}.$$

(8) und (2) liefern aber auch

(9) $$l_k(x) = \int_1^x \log t \, dM_k(t) = \log x\, M_k(x) - \int_1^x \frac{M_k(t)}{t}\, dt \sim M_k(x) \log x.$$

(8) und (9) ergeben die Behauptung (6).

Hilfssatz 6: *Es ist* $(x \to \infty)$

(10) $$\sum_{p \in \mathfrak{T}} \frac{1}{p^{1+\frac{1}{x}} - p^{\frac{k-1}{k}\left(1+\frac{1}{x}\right)} + 1} = \sum_{\substack{p \leq e^x \\ p \in \mathfrak{T}}} \frac{1}{p} - C\tau + S_k + o(1),$$

wobei C die Eulersche Konstante und

(11) $$S_k = \sum_{p \in \mathfrak{T}} \frac{p^{\frac{k-1}{k}} - 1}{p\left(p - p^{\frac{k-1}{k}} + 1\right)}$$

ist.

Beweis: Wirsing [6] beweist

(12) $$\sum_{p \in \mathfrak{T}} \frac{1}{p^{1+\frac{1}{x}}} = \sum_{\substack{p \leq e^x \\ p \in \mathfrak{T}}} \frac{1}{p} - C\tau + o(1) \qquad (x \to \infty);$$

für $x > 0$ ist

$$S_k(x) = \sum_{p \in \mathfrak{T}} \frac{1}{p^{1+\frac{1}{x}} - p^{\frac{k-1}{k}\left(1+\frac{1}{x}\right)} + 1} - \sum_{p \in \mathfrak{T}} \frac{1}{p^{1+\frac{1}{x}}}$$

$$= \sum_{p \in \mathfrak{T}} \frac{p^{\frac{k-1}{k}\left(1+\frac{1}{x}\right)} - 1}{p^{1+\frac{1}{x}}\left[p^{1+\frac{1}{x}} - p^{\frac{k-1}{k}\left(1+\frac{1}{x}\right)} + 1\right]}$$

gleichmäßig konvergent, also

$$\lim_{x \to \infty} S_k(x) = \sum_{p \in \mathfrak{T}} \frac{p^{\frac{k-1}{k}} - 1}{p\left(p - p^{\frac{k-1}{k}} + 1\right)} = S_k . \tag{13}$$

(12) und (13) liefern die Behauptung (10).

Hilfssatz 7 (Hardy-Littlewood [2]): *Es sei $F(t) \geqq 0$ und*

$$f(s) = \int_0^\infty e^{-st} F(t)\, dt$$

für $s > 0$ konvergent. Für $\tau > 0$ gelte

$$f(s) \sim \frac{1}{s^\tau} L\left(\frac{1}{s}\right) \qquad (s \to 0,\ s > 0), \tag{14}$$

wobei $L(x)$ eine für $x > 0$ stetige positive Funktion sei, für die mit jedem $u > 0$

$$\lim_{x \to \infty} \frac{L(ux)}{L(x)} = 1 \tag{15}$$

erfüllt ist. Aus diesen Voraussetzungen folgt

$$\int_0^x F(t)\, dt \sim \frac{1}{\Gamma(1+\tau)} x^\tau L(x) \qquad (x \to \infty) . \tag{16}$$

Vgl. hierzu auch Doetsch [1], S. 511.

§ 2. (Ergebnisse)

Auf Grund von Hilfssatz 5 ist die gesuchte asymptotische Abschätzung von $M_k(x)$ ermittelt, wenn eine solche für $m_k(x)$ angegeben werden kann; dies geschieht im folgenden unter Anwendung von Hilfssatz 7. Dabei benötigen wir noch die Hilfssätze 1, 2 und 6.

Für $s > 0$ setzen wir

$$f(s) = \frac{1}{1+s} \sum_{m \in \mathfrak{M}_k} \frac{1}{m^{\frac{1+s}{k}}} = \frac{1}{1+s} \int_{1-}^\infty \frac{d M_k(t)}{t^{\frac{1+s}{k}}} ,$$

also, nach partieller Integration und Berücksichtigung von (2),

$$f(s) = \frac{1}{k} \int_1^\infty \frac{M_k(t)}{t^{\frac{1+s}{k}+1}}\, dt = \int_0^\infty e^{-st} \frac{M_k(e^{kt})}{e^t}\, dt$$

und

$$F(t) = \frac{M_k(e^{kt})}{e^t} > 0\,.$$

Es ist aber auch

$$(1+s)\,f(s) = \prod_{p\in\mathfrak{T}}\left[1 + \frac{1}{p^{k\left(\frac{1+s}{k}\right)}} + \frac{1}{p^{(k+1)\left(\frac{1+s}{k}\right)}} + \cdots\right] = \prod_{p\in\mathfrak{T}}\left[1 + \frac{1}{p^{1+s} - p^{\frac{k-1}{k}(1+s)}}\right]$$

$$= \prod_{p\in\mathfrak{T}} \frac{1}{1 - \left(p^{1+s} - p^{\frac{k-1}{k}(1+s)} + 1\right)^{-1}} = e^{\sum\limits_{p\in\mathfrak{T}}\sum\limits_{\nu\geqq 1} \frac{1}{\nu\left(p^{1+s} - p^{\frac{k-1}{k}(1+s)} + 1\right)^{\nu}}}\,.$$

Mit

$$L(x) = e^{-\tau\log x + \sum\limits_{p\in\mathfrak{T}}\sum\limits_{\nu\geqq 1} \frac{1}{\nu\left(p^{1+\frac{1}{x}} - p^{\frac{k-1}{k}\left(1+\frac{1}{x}\right)} + 1\right)^{\nu}}} \tag{17}$$

ist also

$$(1+s)\,f(s) = \frac{1}{s^\tau}\,L\!\left(\frac{1}{s}\right), \qquad s > 0\,,$$

woraus (14) folgt.

Um Hilfssatz 7 anwenden zu können, muß noch gezeigt werden, daß für jedes $u > 0$ (15) gilt; ersichtlich genügt es, $u > 1$ zu wählen.

Aus (17) folgt wegen der gleichmäßigen Konvergenz von

$$\sum_{p\in\mathfrak{T}}\sum_{\nu\geqq 2} \frac{1}{\nu\left(p^{1+\frac{1}{x}} - p^{\frac{k-1}{k}\left(1+\frac{1}{x}\right)} + 1\right)^{\nu}}$$

und unter Verwendung von (10) $(x\to\infty)$

$$\log L(x) = -\tau\log x + \sum_{\substack{p\leqq e^x\\ p\in\mathfrak{T}}}\frac{1}{p} - C\tau + S_k + \sum_{p\in\mathfrak{T},\,\nu\geqq 2}\frac{1}{\nu\left(p - p^{\frac{k-1}{k}} + 1\right)^{\nu}} + o(1)\,,$$

also für $u > 1$

$$\log L(ux) - \log L(x) = \sum_{\substack{e^x < p \leqq e^{ux}\\ p\in\mathfrak{T}}}\frac{1}{p} - \tau\log u + o(1) = o(1) \qquad (x\to\infty)\,,$$

letzteres auf Grund von (3). Daraus ergibt sich (15).

Sämtliche Voraussetzungen des Hilfssatzes sind also erfüllt, und wir haben an Stelle von (16)

$$\int_0^{\log x}\frac{M_k(e^{kt})}{e^t}\,dt \sim \frac{1}{\Gamma(1+\tau)}\,(\log x)^\tau\,L(\log x) \qquad (x\to\infty)\,. \tag{18}$$

Hierin erhalten wir für die linke Seite

$$\int_0^{\log x}\frac{M_k(e^{kt})}{e^t}\,dt = \frac{1}{k}\int_1^{x^k}\frac{M_k(t)\,dt}{t^{\frac{k+1}{k}}} = -\int_1^{x^k} M_k(t)\,dt^{-\frac{1}{k}}$$

$$= O(1) + \int_1^{x^k} t^{-\frac{1}{k}}\,dM_k(t) = O(1) + m_k(x^k) \qquad (x\to\infty)\,, \tag{19}$$

wieder unter Verwendung von (2).

Für die rechte Seite von (18) bekommen wir

$$\frac{1}{\Gamma(1+\tau)}(\log x)^{\tau} L(\log x) = \frac{c_4}{\Gamma(1+\tau)} \cdot e^{\sum\limits_{\substack{p \leq x \\ p \in \mathfrak{T}}} \frac{1}{p} + o(1)}$$

mit

$$c_4 = e^{S_k - C\tau + \sum\limits_{p \in \mathfrak{T}, \nu \geq 2} \frac{1}{\nu\left(p - p^{\frac{k-1}{k}} + 1\right)^{\nu}}}.$$

Dies und (19) in (18) eingesetzt, ergibt mit $x^{\frac{1}{k}}$ an Stelle von x

$$m_k(x) \sim \frac{c_4}{\Gamma(1+\tau)} \cdot e^{\sum\limits_{\substack{p \leq x^{\frac{1}{k}} \\ p \in \mathfrak{T}}} \frac{1}{p}} \qquad (x \to \infty).$$

Hilfssatz 5 liefert nun den

Satz 1: *$\mathfrak{M}_k$ sei die Menge aller natürlichen Zahlen, deren Primteiler sämtlich*
a) in mindestens k-ter Potenz ($k \geq 1$) *auftreten und*
b) aus einer Menge $\mathfrak{T}$ von Primzahlen stammen, für die (1) *mit einem* $\tau > 0$ *gilt.*
Es sei S_k die durch (11) *definierte Konstante,*

$$B_k = \sum_{p \in \mathfrak{T}, \nu \geq 2} \frac{1}{\nu\left(p - p^{\frac{k-1}{k}} + 1\right)^{\nu}}$$

und C die Eulersche Konstante.
Dann gilt für die Anzahlfunktion von $\mathfrak{M}_k$

$$M_k(x) \sim k \frac{e^{B_k + S_k - C\tau}}{\Gamma(\tau)} \cdot \frac{x^{\frac{1}{k}}}{\log x} \cdot e^{\sum\limits_{\substack{p \leq x^{\frac{1}{k}} \\ p \in \mathfrak{T}}} \frac{1}{p}} \qquad (x \to \infty). \tag{20}$$

Für $k = 1$ ist $S_k = S_1 = 0$, und es folgt

Satz 2: (WIRSING [6]):

$$M_1(x) \sim \frac{1}{\Gamma(\tau)} e^{-C\tau + \sum\limits_{p \in \mathfrak{T}, \nu \geq 2} \frac{1}{\nu p^{\nu}}} \cdot \frac{x}{\log x} \cdot e^{\sum\limits_{\substack{p \leq x \\ p \in \mathfrak{T}}} \frac{1}{p}}.$$

Ist andererseits $\mathfrak{T}$ die volle Primzahlmenge $\mathfrak{P}$, also $\tau = 1$, so erhält man für die eingangs eingeführte Menge $\mathfrak{A}_k$ den

Satz 3: *$\mathfrak{A}_k$ sei die Menge aller natürlichen Zahlen, deren Primteiler alle in mindestens k-ter Potenz* ($k \geq 1$) *auftreten. Es sei*

$$S_k^* = \sum_{p \in \mathfrak{P}} \frac{p^{\frac{k-1}{k}} - 1}{p\left(p - p^{\frac{k-1}{k}} + 1\right)},$$

$$B_k^* = \sum_{p \in \mathfrak{P}, \nu \geq 2} \frac{1}{\nu\left(p - p^{\frac{k-1}{k}} + 1\right)^{\nu}}.$$

Dann gilt für die Anzahlfunktion $A_k(x)$ von $\mathfrak{A}_k$

$$A_k(x) \sim e^{B_k^* - B_1^* + S_k^*} \cdot x^{\frac{1}{k}}.$$

Beweis: Bekanntlich ist

$$\sum_{\substack{p \leq x \\ p \in \mathfrak{P}}} \frac{1}{p} = \log\log x + C - B_1^* + o(1) \qquad (x \to \infty)\,; \tag{21}$$

also hat man nach (20)

$$A_k(x) \sim k e^{B_k^* + S_k^* - C} \cdot e^{C - B_1^*} \cdot \frac{x^{\frac{1}{k}}}{\log x} \log x^{\frac{1}{k}} = e^{B_k^* - B_1^* + S_k^*} \cdot x^{\frac{1}{k}}.$$

Bemerkung: Notwendig und hinreichend dafür, daß

$$M_k(x) \sim c_5 x^{\frac{1}{k}}$$

wird, ist das Bestehen einer Relation ($\tau = 1$)

$$\sum_{\substack{p \leq x \\ p \in \mathfrak{T}}} \frac{1}{p} = \log\log x + c_6 + o(1) \qquad (x \to \infty)\,. \tag{21*}$$

Ist $\mathfrak{T}$ die Menge aller Primzahlen einer primen Restklasse mod n, so wird statt (21)

$$\sum_{\substack{p \leq x \\ p \in \mathfrak{T}}} \frac{1}{p} = \frac{1}{\varphi(n)} \log\log x + c_7 + o(1) \qquad (x \to \infty) \tag{22}$$

(vgl. auch Landau [3], § 110); enthält $\mathfrak{T}$ die Primzahlen mehrerer, etwa r, solcher primen Restklassen mod n, so ist entsprechend ($x \to \infty$)

$$\sum_{\substack{p \leq x \\ p \in \mathfrak{T}}} \frac{1}{p} = \tau \log\log x + c_8 + o(1) \qquad \left(\tau = \frac{r}{\varphi(n)}\right). \tag{23}$$

Satz 1 liefert dann nach Einsetzen von (22) bzw. (23) in (20)

$$M_k(x) \sim c_9(k) \frac{x^{\frac{1}{k}}}{(\log x)^{1-\tau}}\,.$$

Die weitere Spezialisierung $k = 1$ führt auf einen bekannten Satz von Landau [4, 5].

Literatur

[1] Doetsch, G.: Handbuch der Laplace-Transformation I. (1950). — [2] Hardy, G. H. and J. E. Littlewood: Notes on the theory of series. XI. On Tauberian theorems. Proc. Math. Soc. London (2) **30**, 23—37 (1929). — [3] Landau, E.: Handbuch der Lehre von der Verteilung der Primzahlen I (1909). — [4] Landau, E.: Handbuch der Lehre von der Verteilung der Primzahlen II (1909). — [5] Landau, E.: Lösung des Lehmerschen Problems. Amer. J. Mathematics **31**, 86—102 (1909). — [6] Wirsing, E.: Über die Zahlen, deren Primteiler einer gegebenen Menge angehören. Arch. Math. VII, 263—272 (1956).

(Eingegangen am 19. März 1959)